Developing a Triangular Tessellation Method for the Analysis of Medium Ring Pucker Conformations

A thesis submitted to the

UNIVERSITY OF CAPE TOWN

in fulfillment of the requirements for the degree of

MASTER of SCIENCE

By

Pegah Khalili

BSc. (Hons) (University of Kwazulu Natal - Pietermaritzburg)

May, 2013

Supervisor: Professor Kevin J. Naidoo

Developing a Triangular Tessellation Method for the Analysis of Medium Ring Pucker Conformations

Pegah Khalili **May 2013**

ABSTRACT

The main focus of this thesis is to investigate the relative conformational flexibilities of α-, β- and γ-cyclodextrins in water by analysing their *macrocyclic* ring puckering motion from Molecular Dynamics (MD) simulations. In particular, the puckering of the CDs is investigated through a ***coarse grained analysis of full atomistic simulations***, where the CD conformational motions are studied on the *macrocyclic* scale rather than the *atomistic* scale. The flexibilities of the cyclodextrins (CDs) are then compared to their experimentally-observed aqueous solubility trend in order to try explain the anomalously low solubility of β-cyclodextrin. β-CD has important applications in industry, such as the pharmaceutical industry, thus exploring the conformational reasons for its low solubility can help to design more effective cyclodextrin-based products in future.

The ring puckering of the CDs is measured quantitatively using a reduced system of puckering coordinates based on the method of triangular tessellation. The triangular tessellation definition for *monocylic* 6-membered rings is first extended to 7- and 8-membered rings, and the corresponding puckering coordinates are derived mathematically. The *macrocyclic* CD rings are then simplified to monocyclic representations through an appropriate coarse graining of the molecules (specifically, α-, β- and γ-cyclodextrins are simplified to 6-, 7- and 8-sided rings, respectively), and the corresponding triangular tessellation definition is then used to measure their macrocyclic puckering. The rates of decay of the puckering motion are then calculated using *time correlation functions*, from which the relative flexibilities of the CDs is determined. Probability distributions are also used to investigate the ranges of the CD puckering. In addition, the horizontal contraction and expansion of the macrocyclic rings (termed "breathing" herein) is analysed to supplement the puckering analysis.

Puckering coordinates based on the triangular tessellation of 6-membered rings have been used previously to characterise all 38 canonical states of cyclohexane. In this thesis, a *systematic procedure* is developed to generate the triangular tessellation puckering coordinates of all the canonical states of 6-, 7- and 8-membered rings, and the coordinates for all canonical states of cycloheptane and cyclooctane are subsequently generated. These puckering coordinates can be useful not only in the conformational analysis of cyclohexane, cycloheptane and cyclooctane, but also to quantitatively characterise the conformations of 6-, 7- and 8-membered rings in general, both from experimental and computational studies.

Acknowledgements

I would like to thank my supervisor, Prof. Naidoo, for his enthusiasm and support during the course of this project.

I would also like to thank several other people:

- Chris, Krishna, Riedaa, Werner, and Jestin for their help during various stages of the project and write-up – I really appreciate your kindness and support during my stay at UCT.
- Annalisa, Richard, Ranga, and Kyle, for their help especially during the beginning stages of the project.
- Dr. Venter, for helping with improving qualities of some pictures in this thesis.
- All other SCRU members, for all their support and friendship.
- Louise, for her constant support, friendship and caring nature during my stay in Cape Town – thank you for all your kindness during my stay here.
- To my family, for their love and support during these past years.

I would also like to acknowledge SARChI and the University of Cape Town for providing funding.

Presentations and Publications

Parts of the work done in this project have been presented on a poster entitled "Analysis of Ring Conformations" at the Frontiers in Scientific Computing symposium, held on 1-2 October 2012 in Cape Town, South Africa.

The following paper (on which Chapter 5 is based) has been published from the work carried out in this thesis:
"**Interpreting medium ring canonical conformers by a triangular plane tessellation of the macrocycle"** J. Chem. Phys. **138**, 184110 (2013); http://dx.doi.org/10.1063/1.4803698

List of Abbreviations

CD	Cyclodextrin
α-CD	Alpha cyclodextrin
β-CD	Beta cyclodetrin
γ-CD	Gamma cyclodextrin
Å	Angstrom (10^{-10} m)
fs	Femtosecond (10^{-15} s)
ps	Picosecond (10^{-12} s)
ns	Nanosecond (10^{-9} s)
QM	Quantum Mechanics
MM	Molecular Mechanics
MD	Molecular Dynamics
NMR	Nuclear Magnetic Resonance
IR	InfraRed (Spectroscopy)
TCF	Time Correlation Function
RMSD	Root Mean Square Deviation
TT	Triangular Tessellation
APC	Angular Puckering Coordinates
2D	Two-dimensional
3D	Three-dimensional
TS	Transition State
VMD	Visual Molecular Dynamics
CHARMM	Chemistry at Harvard Macromolecular Mechanics
CSFF	Carbohydrate Solution Force Field
HGFB	Ha, Giammona, Field and Brady carbohydrate force field
PHLB	Palma, Himmel, Liang and Brady carbohydrate force field
PBC	Periodic Boundary Conditions
IUPAC	International Union of Pure and Applied Chemistry
IUB	International Union of Biochemistry

* Other abbreviations that are not listed here will be explained in the thesis chapters.

Table of Contents

Index of Figures and Tables

CHAPTER 1

CHAPTER 2

CHAPTER 3

CHAPTER 4

CHAPTER 5

CHAPTER 6

APPENDIX A

APPENDIX B

APPENDIX C

APPENDIX D

Chapter 1

Introduction

Cyclic molecules play important roles in several chemical and biological processes – for example, carbohydrates are a broad class of cyclic molecules that are important in activities of the biological cell, such as energy transport and cell recognition [1]. A better understanding of the reactions and processes involving cyclic molecules (hereafter also referred to as *rings*) can be achieved through studying the structures of these molecules as well as their conformational behaviour.

One class of cyclic molecules are the cyclodextrins (macrocyclic carbohydrate molecules composed of varying numbers of glucose monomer units), which are important in a variety of industrial applications [2-7]. Beta-cyclodextrin (composed of seven glucose units) is particularly useful in the pharmaceutical industry as a drug delivery system [8], but its inhibiting property is its very low aqueous solubility in comparison to alpha-cyclodextrin (α-CD) and gamma-cyclodextrin (γ-CD), which are composed of six and eight glucose units, respectively – the solubility trend of these cyclodextrins (CDs), in decreasing order, is: γ-CD > α-CD ≫ β-CD [2]. It is important to investigate the underlying molecular reasons for the anomalous solubility of beta-cyclodextrin (β-CD) in relation to α- and γ-CD, as this may help to design more effective and efficient drug-delivery systems and other industrial products in future. In a previous study by Naidoo *et al.* [8], dynamic simulations were used to investigate the aqueous solubility trend of α-, β- and γ-CD in relation to their conformational flexibility, by measuring the time correlation of the cyclodextrin (CD) monomeric motions (angular tilting and vertical librations of the glucose units) and macrocyclic motions (glycosidic bond rotations in the molecule) – this study demonstrates a useful correlation between the flexibilities and solubilities of these three particular CDs, showing that as the flexibility of the CD increases, so does its solubility. Part of this thesis follows on from this study, but here, in order to explain the solubility trend of α-, β- and γ-CD, a newly-defined metric is used to measure their conformational flexibilities, namely, the CD macrocyclic ring deformations (or macrocyclic ring ***puckering***). In particular, dynamic simulations are used to measure the CD puckering, and this motion is then correlated over time to gain

insight into the CD flexibilities. The deformations (or conformational changes) in ring systems play significant roles in their physical and chemical properties (as will be explained in Section 1.1), which is why the ring puckering motions of α-, β- and γ-CD are being analysed here, to gain insight into their solubility trend and the anomalous solubility of β-CD (where solubility is a physical property of the system). In addition, the correlation between flexibility (as measured by ring puckering) and solubility of these three CDs is being investigated in this work to establish if the results are consistent with those of reference [8]. To supplement the puckering analysis, the planar deformation motion of the CD rings, termed "breathing" herein, will be defined and investigated.

The puckering of the CDs is studied here through a ***coarse grained analysis of full atomistic simulations***, that is, the CD conformational motions are investigated on the *macrocyclic* scale rather than on the *atomistic* scale. To measure the CD macrocyclic puckering, a quantitative model is first chosen to measure the changes in the ring conformations at different times during the simulations. In this thesis, the quantitative model of *triangular decomposition* [9] is selected (for reasons discussed in Section 1.2) – the model for 6-membered *monocyclic* rings [9] is further developed for 7- and 8-membered rings, and then used to model the *macrocyclic* CDs through an appropriate coarse graining of the rings. In this work, the term ***triangular tessellation* (TT)** is used to describe the model rather than the term *triangular decomposition*, and the puckering coordinates based on the triangular tessellation of the ring will hereafter be referred to as ***angular puckering coordinates* (APC)**.

In addition to the quantitative description of ring systems in dynamics simulations, it is also useful to be able to characterise conformations *qualitatively* in terms of canonical conformations (e.g. *chair*, *boat*), which may be more chemically meaningful, and also to be able to relate the quantitative characterisation to the qualitative. Thus in addition, part of this thesis develops a general, systematic method to generate the APC of the canonical conformers (characteristic shapes) of an *N*-membered ring (for $N = 6, 7, 8$), which is subsequently used to generate the APC of all the canonical conformers of cycloheptane and cyclooctane – this can be useful in future dynamics studies of 7- and 8-membered rings, to relate the qualitative and quantitative descriptions.

The significance of ring conformational analysis in general is described in Section 1.1, followed by a brief description of ring puckering models in Section 1.2. In addition, the reasons for selecting the TT model to study the CDs in this thesis are presented in Section 1.2. Section 1.3 presents an overview of the structure, applications and conformational behaviour of CDs, with emphasis on α-, β- and γ-CD. Finally, the objectives of this work are given in Section 1.4.

1.1 Conformational Analysis of Rings

The *conformation* of a molecule can be defined, in part, as the spatial arrangement adopted by the atoms as a result of torsional motions about the single bonds in the molecule [10]. There are also other conformational changes a ring can undergo, including inversion at an atom, ring inversion (flipping), ring pseudorotation or bond pseudorotation. In addition, rotations are also possible in rings with double bonds with weak pi-bonding. *Conformational analysis* then refers to the study of different conformations of the molecule to help explain a system's observed physical and chemical behaviour [10]. The conformational changes of rings are an important aspect to consider in the chemical reactivity, physical properties and biological function of a system [11]. For example, the process of DNA duplication depends on the conformational changes associated with carbohydrate rings [12], and the elasticity of polysaccharides (long-chain carbohydrates) such as amylose, dextran and pullulan is due to changes between the *chair* and *boat* conformations of their component pyranose rings [13, 14]. The importance of conformational analysis in ring systems with regards to physical properties and chemical reactivity is discussed in more detail below – only ***non-aromatic*** rings are considered in this thesis.

1.1.1 Physical Properties

A notable case where the conformational changes of ring systems influence their physical properties is in the area of host-guest chemistry. There are several organic macrocyclic molecules that can act as ligands (hosts) and bind different ionic and molecular species (guests) in their central cavity to form inclusion complexes [15]. Macrocyclic ligands include compounds such as porphyrins, corrins, crown-ethers and cyclodextrins [15] – **Figure 1.1** depicts a few simple examples of these compounds.

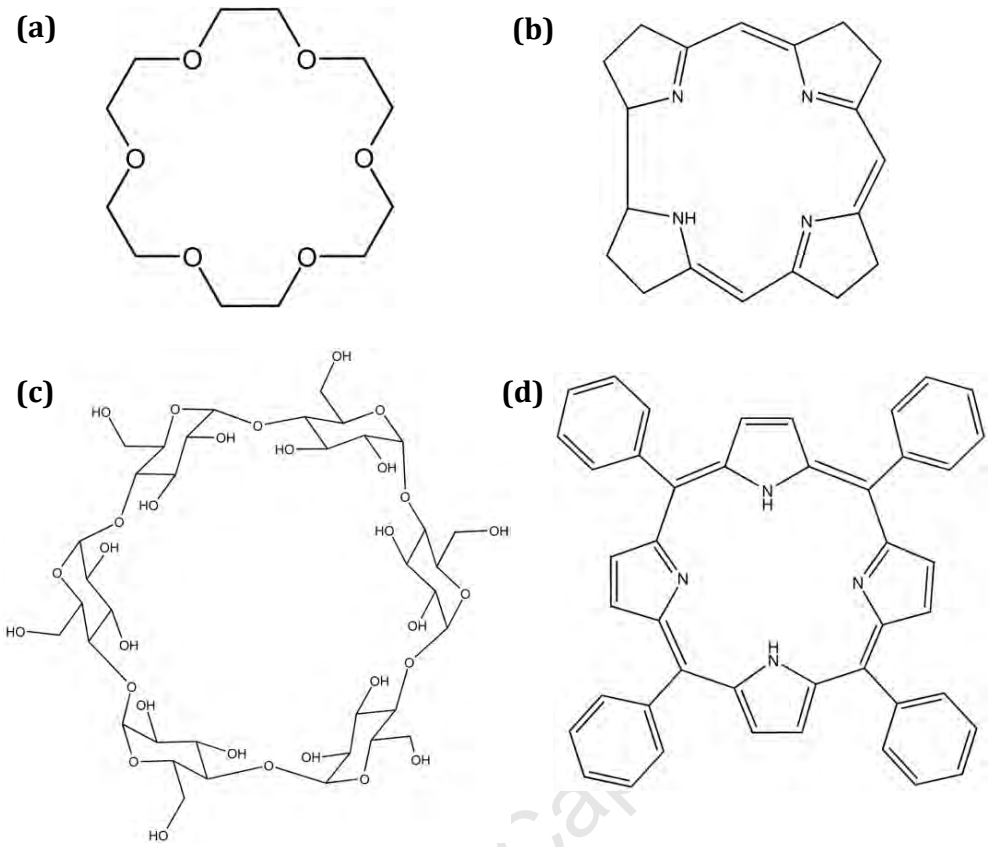


Figure 1.1: Structures of **(a)** 1,4,7,10,13,16-hexaoxacyclooctadecane (a crown ether), **(b)** corrin, **(c)** cyclomaltohexaose (α-cyclodextrin), and **(d)***meso*-Tetraphenylporphyrin.

The complexes formed with macrocyclic ligands are important in many biological processes such as photosynthesis, oxygen transport in mammals, selective transportation of ions across membranes, catalysis of chemical processes, and enzyme-substrate interactions, leading to several medical and industrial applications [15]. Several factors play a role in the complexation process, such as intermolecular interactions, solvation effects, cavity-size, as well as the conformational changes (or distortions) the host undergoes in order to incorporate the guest [15]. The conformational *changes* are often required for good complexation (for example, they can optimise a particular host-guest interaction), and the conformational *flexibility* of the host also plays a role in its complexing behaviour (how and what species it can complex) [15]. This illustrates the importance of conformational analysis in complexation processes of macrocycles.

The conformational arrangement of the macrocyclic ligand around the guest can also affect the *stability* of the complex. An example includes the crown ether 1,4,7,10,13,16-

hexaoxacyclooctadecane (18-crown-6), depicted in **Figure 1.1(a)**. 18-crown-6 can complex both K^+ and Na^+ ions, but adopts different geometries (or conformations) around these ions, as revealed by X-ray structures of its potassium thiocyanate (KNCS) [16] and sodium thiocyanate (NaNCS) [17] complexes. In the case of the KNCS complex, the ionic radius of K^+ matches the approximate radius of the macrocyclic cavity of 18-crown-6 well, so that the six oxygens of the crown ether arrange in an almost planar and unstrained manner around the K^+ ion. However, in the NaNCS complex, the crown ether bends into an irregular (i.e. distorted) conformation to accommodate the Na^+ ion, where one ether oxygen lies out of the mean plane of the other five oxygens. The KNCS complex has also been reported to be more stable than the NaNCS complex, and the different conformations of the crown ether in these complexes appears to explain this difference in stability [15].

The observed *solubilities* of macrocyclic ligands such as α-, β- and γ-CD have also been studied in relation to their conformational behaviour. For example, through computational studies of their macrocyclic and monomeric motions in water [8], the relative conformational flexibilities of these three CDs has been related to their experimentally-observed solubility trend, i.e. as the CD flexibility increases, its aqueous solubility also increases.

1.1.2 Chemical Reactivity

Although this thesis does not deal with chemical reactions, the significance of ring conformational changes in chemical reactivity is briefly outlined below, to provide a more complete background on the scope of conformational analysis.

Ring conformational changes form an important part of several chemical and biological processes, such as enzyme-catalysed reactions. One such process is the hydrolysis of glycosidic bonds in carbohydrates, which is catalysed by glycosyl hydrolases (or glycosidases) [18-20]. During the reaction, the carbohydrate binds to the enzyme and undergoes different distortions (or deformations) to reach the final product [21, 22]. In several glycosidase complexes for example, the relevant pyranose unit of the carbohydrate binds at the active site of the enzyme and adopts a distorted *boat* or *skew* conformation, instead of the more favourable $^4C_1$ *chair* conformation [22]. These kinds

of distortions place the relevant regions of the carbohydrate moiety in favourable positions relative to the catalytic enzyme residues, thereby facilitating the cleavage of the glycosidic bond [22]. The carbohydrate conformation in the transition state of several glycosidases also determines the reaction pathway that is followed, leading to the formation of specific products [21].

In order to better understand the mechanistic details of glycosidases, as well as of enzymes in general, studies have been carried out on the conformational itineraries of both isolated ring systems [14, 21, 22] and enzyme substrates [23, 24]. In a recent computational study for example, the Free Energies from Adaptive Reaction Coordinate Forces (FEARCF) method was used to sample the many conformations of a β-D-glucose ring and to plot its three-dimensional (3D) conformational free energy surface [21]. From this surface, the free energy pathways connecting the low energy conformations were located and plotted, showing the locations, relative energies and conformations of the transition states and local minima – these pathways can be a useful tool in elucidating the mechanisms of enzymes such as glycosidases [21].

The conformation of the substrate in the *transition state* (TS) of an enzyme-catalysed reaction is particularly important, not only to help understand the catalytic procedure, but also to aid in designing transition state inhibitors (TS inhibitors) [25]. TS inhibitors mimic the structure of the substrate TS [26], but also have particular feature/s that allow them to inhibit (or block) the action of the enzyme. TS inhibitors can be used to develop novel therapeutic agents to treat human diseases – for example, inhibitors of glycosidases can potentially be used to treat viral infections such as HIV and influenza, and also to treat diseases such as diabetes, cancer and lysosomal storage disorders [26, 27]. As a result, several studies have been carried out into the structures and binding of TS inhibitors [25, 26]. A notable example involves the enzyme α-amylase (α-1,4-glucan-4-glucanohydrolase), which is found in the saliva and pancreatic secretions of humans [27]. Salivary α-amylase catalyses the hydrolysis of α-(1,4) glycosidic bonds in starch to form smaller oligomers which pass into the gut, then pancreatic α-amylase breaks down these oligomers into even smaller oligosaccharides, and finally α-glucosidases break down the oligosaccharides into glucose, which then passes into the blood stream [27]. Inhibitors of pancreatic α-amylase and α-glucosidases have been developed, which slow

the digestion of the starch and oligosaccharides [27], thereby controlling blood sugar levels. For example, the naturally-occuring inhibitor *acarbose* is a strong inhibitor of human pancreatic α-amylase, and has been studied in the treatment of diabetes – it is reported that the structure of acarbose likely mimics the flattened conformation of glucose in the TS of the amylase during starch hydrolysis, which contributes to its strong inhibition of the enzyme [27].

1.2 Puckering Models

The different ring conformations of a molecule are also referred to as *puckers*, and the process of changing conformations due to out-of-plane bending motions (i.e. torsional movements) is called *puckering*. Although there are a vast number of conformations for a ring, these can be divided into a relatively small number of *characteristic* conformations and a vast number of intermediate forms. The characteristic forms (also known as the *canonical* conformations) can be classified further into different *states*, which describe the positions of the ring atoms relative to a suitably defined plane. Systematic techniques have been developed to characterise ring puckers both *qualitatively* and *quantitatively*. The *qualitative* nomenclature of puckers is *discrete* in the sense that it is applied to the naming of the finite number of canonical forms and their states, while their *quantitative* characterisation involves developing *continuous* models that can be applied to both canonical and intermediate forms. In quantitative models, a reduced number of parameters are defined based on perpendicular displacements from a mean plane of the ring [28-31], the intracyclic torsion angles [11, 32-37], or the triangular tessellation (TT) of the ring [9, 28, 38, 39]. The characterisation of ring pucker will be discussed in more detail in Chapter 4, with particular emphasis on quantitative models.

In this thesis, the quantitative model of TT is used to study the puckering of the macrocyclic rings of α-, β- and γ-CD in water. Although the TT method has been defined for monocyclic rings, it can also be applied to macrocyclic rings by using an appropriate coarse graining of the macrocycle (i.e. simplifying the macrocyclic ring to a monocyclic representation). In brief, TT involves decomposing a ring into triangular planes which rotate relative to each other as the ring takes on different conformations – the angles formed between the triangular planes then form the angular puckering coordinate set

describing a particular conformation. For example, a 6-membered ring conformation can be described by three APC.

There are different points to consider with regards to why the TT method was chosen to describe the macrocyclic puckering of the CD rings in this thesis:

1. APC based on the TT model present an intuitive way to consider ring puckering of monocyclic rings, and can be extended to describe macrocyclic rings relatively easily, through an appropriate coarse graining of the macrocycle.
2. In this thesis, the *dynamic* motion of α-, β- and γ-CD is measured from Molecular Dynamics simulations. Since the CD molecules are *homogeneous* rings, composed of identical glucose monomer units (as discussed in Section 1.3), the dynamic conformational motion across the CD molecule can be expected to be similar, on average, if a long enough simulation time is chosen.
3. The APC are all the same *type* of quantity, i.e. *angular* quantities.

Due to points 2. and 3. above, the *correlation* of the individual APC *over time* (as measured by time correlation functions in this work) will therefore follow a similar pattern and rate of decay. In this way, the correlations of the different APC for each CD can be averaged to yield an *average correlation* at each time during the simulation. The rates of decay of these average correlations for α-, β- and γ-CD can then be compared to gain insight into their relative flexibilities.

1.3 Cyclodextrins

1.3.1 Structure and Properties

CDs are cyclic oligosaccharides composed of different numbers of glucose subunits, and are produced from the degradation of starch by glucosyltransferase enzymes [40]. Due to their cyclic structure, CDs have an internal cavity which can incorporate guest molecules and form inclusion complexes, leading to several industrial applications of CDs. The three main CDs are α-cyclodextrin (α-CD, cyclomaltohexaose), β-cyclodextrin (β-CD, cyclomaltoheptaose) and γ-cyclodextrin (γ-CD, cyclomaltooctaose) which are composed of six, seven and eight glucose units, respectively [2, 40], and are illustrated in **Figure 1.2**.

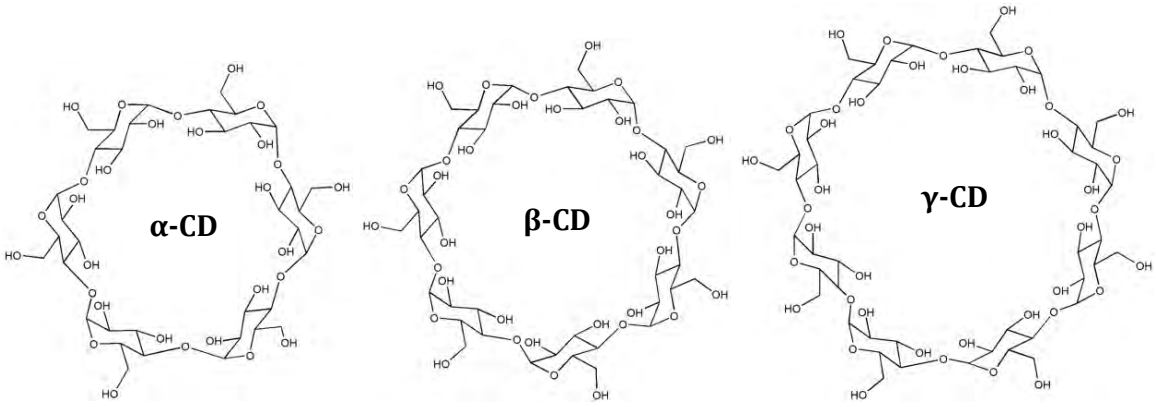


Figure 1.2: Schematic diagram of the structures of α-, β- and γ-cyclodextrin.

CDs are composed of α-D-glucose units [41], where the hydroxyl group on the anomeric carbon (C1) is in the axial position (as opposed to β-D-glucose where the hydroxyl on C1 is in the equatorial position). The individual glucose units are then linked together through α-(1,4) glycosidic bonds [2, 40], where C1 of one glucose unit is bonded to O4 of the adjacent unit. The glucose units adopt the $^4C_1$ conformation [2, 40], where C4 is positioned above the plane defined by C2, C3, C5 and O5, while C1 is positioned below this plane. The structure of α-D-glucose in the $^4C_1$ conformation (together with the ring numbering scheme) is depicted in **Figure 1.3(a)**, and the α-(1,4) glycosidic bond for a portion of a CD molecule is shown in **Figure 1.3(b)**.

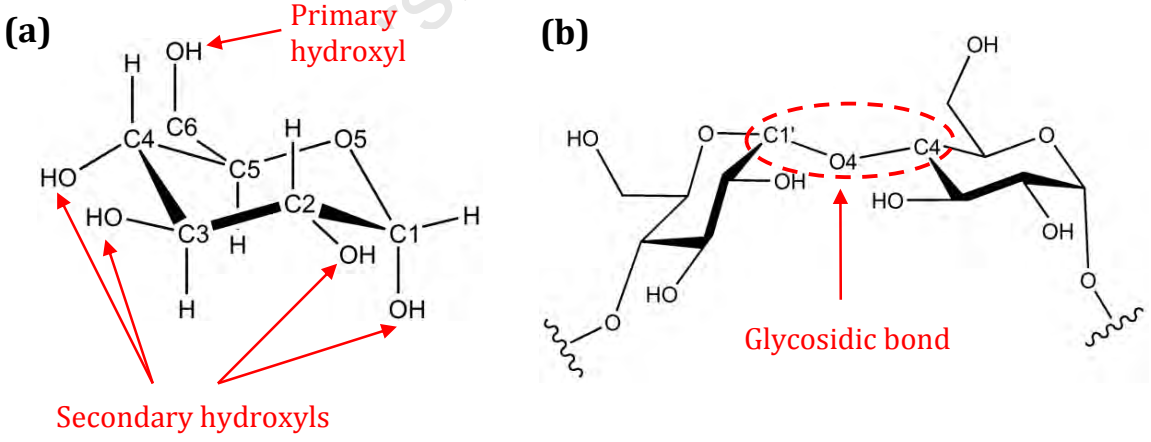


Figure 1.3: **(a)** Structure of α-D-glucose in the $^4C_1$ conformation, showing the ring numbering scheme and positions of primary and secondary hydroxyl groups. **(b)** A portion of a CD, showing the α-(1,4) glycosidic bond between adjacent glucose units. The atoms C4 and O4 are part of one glucose unit, and C1' belongs to the adjacent unit.

According to early X-ray studies, CDs were considered to be highly symmetrical, *rigid* structures – in the case of α-, β- and γ-CD, the rings would have C_6, C_7 and C_8 symmetry, respectively, with the glycosidic oxygens lying in a plane at the vertices of a hexa-, hepta- and octagon, respectively [42]. However, the idea of rigidity is not consistent with the ability of CDs to easily complex guest molecules of different shapes [42].

Different experimental (e.g. NMR) and computational studies have shown that CDs are rather flexible molecules, both in solution and the solid state [42]. In solution, the flexibility of CDs arises from motions on the monomeric scale (involving the glucose units) and on the macrocyclic scale (involving the macrocyclic ring *as a whole*). Examples of monomeric motions are horizontal tilting and vertical librations of the glucose units from the macrocyclic plane [8], while rotations about the glycosidic bonds C1'–O4 and O4–C4 in **Figure 1.3(b)** are classified as macrocyclic motions [8, 43-45]. A more detailed summary of the conformational behaviour of CDs is given in Chapter 6.

The 3D structure of CDs is also an important factor to consider, as it plays a role in their host-guest chemistry. In the case of α-, β- and γ-CD, the rings have the shape of a truncated cone (or torus) enclosing the macrocyclic cavity, as depicted in **Figure 1.4**.

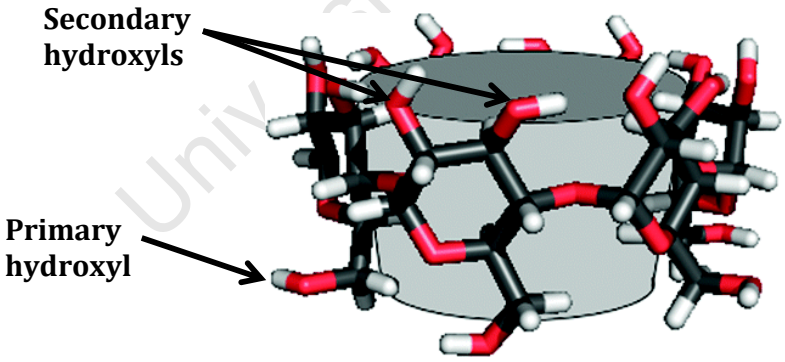


Figure 1.4: The truncated cone structure of α-, β- and γ-CD, shown here for α-CD and taken from reference [45]. Secondary hydroxyl groups are positioned on the wider, upper rim of the CD ring, while primary hydroxyls are on the narrower, lower rim.

Due to the $^4C_1$ conformation of the glucose units, all the secondary hydroxyl (OH) groups (i.e. those attached to C2 and C3) are located on one of the two edges of the macrocyclic CD ring, while the primary OH groups (i.e. those attached to C6) are positioned on the other edge. More specifically, the secondary hydroxyls are positioned on the wider rim

of the ring, while the primary hydroxyls are on the narrower rim [2, 40], as depicted in **Figure 1.4**. The inside of the CD cavity is lined with the hydrogen atoms on C3, C5 and C6, rendering the environment of the cavity *hydrophobic* [40], thus allowing the CD to complex hydrophobic molecules (such as organic moieties) within the cavity. On the other hand, the hydroxyl groups on the exterior rims of the CD are *hydrophilic* and therefore allow the CD molecule (or CD complex) as a whole to be soluble in water [40]. The solubility of organic molecules in aqueous media can therefore be increased by complexation with these CDs, which leads to several useful applications – for example, CDs can increase the aqueous solubility of hydrophobic pharmaceutical drugs [6] for human consumption (the human body being largely aqueous in nature). A further characteristic of the CD cavity is that the lone pairs of the glycosidic oxygens (O4) point towards the inside of the cavity, causing a region of high electron density and imparting to the CD some Lewis base character [2]. **Table 1.1** summarises some of the important characteristics and structural properties of these CDs.

Table 1.1: Some important characteristics and structural properties of α-, β- and γ-CD, taken from reference [2].

	α-CD	**β-CD**	**γ-CD**
Number of glucose units	6	7	8
Molecular weight (g/mol)	972	1135	1297
Aqueous solubility at 25°C (g/100ml)	14.5	1.85	23.2
Cavity diameter (Å)	4.7 – 5.3	6.0 – 6.5	7.5 – 8.3
Height of torus (Å)	7.9 ± 0.1	7.9 ± 0.1	7.9 ± 0.1
Approx. volume of cavity ($Å^3$)	174	262	427
Crystal water (wt %)	10.2	13.2 – 14.5	8.13 – 17.7
Diffusion constant at 40°C	3.443	3.224	3.000

CD complexes have applications in a variety of areas such as catalysis, separation technology and medicine [46]. It is therefore important to investigate their structure and the complexation process in order to gain a better understanding of their behaviour. Different techniques can be used to study CDs and their complexes, such as X-ray crystallography [40, 47], NMR spectroscopy [46], and computational methods [41]. X-ray crystallography is suited to studying solid-state structures, while NMR and

computational techniques can be used to analyse the structure and dynamics of CD complexes in solution. Computational methods in particular, can simulate the behaviour of CDs on the atomic scale, providing a detailed picture of various interactions in the CD system. Thermodynamic quantities of CD complexation reactions have also been measured by various experimental techniques, such as UV-vis spectroscopy, circular dichroism, fluorescence and chromatography [48].

1.3.2 Anomalous Solubility of β-cyclodextrin

In order to form inclusion complexes, the guest molecule must fit into cavity of the CD. With regards to α-, β- and γ-CD, aliphatic molecules complex best with α-CD, small aromatic compounds complex well with β-CD, and larger aromatic species are better suited to γ-CD [8]. Several pharmaceutical drugs consist of aromatic moieties of low molecular weight that are too big to fit into the cavity of α-CD, but fit into the β-CD and γ-CD cavities [8]. However, γ-CD is comparatively very expensive – the prices of α-, β- and γ-CD in Rands per gram (R/g) are R251.21/g, R30.78/g and R2760.00, respectively (as obtained from the Sigma-Aldrich catalogue on 5 December 2012). Thus with regards to drug delivery, β-CD is the preferred choice of the three CDs [8]. However, an inhibiting property of β-CD is its unusually very low aqueous solubility as compared to the other two CDs (see **Table 1.1**), i.e. the solubility of β-CD is much lower than that of α-CD, whereas one might expect its solubility to be intermediate between those of α- and γ-CD.

Attempts have been made to increase the water solubility of CDs by chemical modification of some or all of the hydroxyl groups on C2, C3 and C6 of the glucose monomer units. For example, substituting the hydroxyls with neutral groups (e.g. hydroxypropyl groups) or ionic groups (e.g. carboxymethyl, tertiary amine or quaternary amine groups) causes the aqueous solubility of the modified CD to increase by 60% or higher, as compared to the original CD [3]. Another example involves amorphous mixtures of hydroxyalkylated CDs, where the varying degrees of substitution on the CD molecules inhibits their crystallisation and thereby results in the hydroxyalkylated CDs having a higher water solubility than the unsubstituted CDs [4]. In another study, Shikata *et al.* [49] showed that as the degree of methylation on α-, β- and γ-CD increased (i.e. as more hydroxyls were replaced by methoxy groups), the

interaction with surrounding water molecules increased, thereby increasing the aqueous solubilities of the modified CDs. However, although an increase in the extent of hydroxyl substitution can increase the water solubility, the substituted groups increase the steric hindrance around the CD host, thereby decreasing the complexing abilities of the CD [4]. Therefore, the degree of hydroxyl substitution has to be chosen carefully to preserve the complexing behaviour of the CD.

It is also important to investigate the underlying molecular reasons for the anomalous solubility of β-CD as compared to α- and γ-CD, which has led to different experimental and computational studies on the structure and internal dynamics of these CDs [8, 50-52]. Computer simulation studies of α-, β- and γ-CD in water have shown that the macrocyclic motions of β-CD are more restricted than the other two CDs, and that their relative conformational flexibilities increase in the order β-CD < α-CD < γ-CD [8] – this order mirrors the solubility trend and supports the idea that the low solubility of β-CD is linked to its decreased conformational flexibility. Another factor to consider is the extent of intramolecular hydrogen bonding in the CDs (occurring between the hydroxyls on C2 and C3 of adjacent glucose units). Different studies on CDs and their derivatives suggest a trend, at least for the particular CDs studied, that as the intramolecular hydrogen bonding decreases, the flexibility of the CD increases [8, 53]. The relation to solubility has also been studied. In comparing α-CD to permethylated α-CD (pmα-CD) for example, all the OH groups on α-CD are changed to methoxy (OCH_3) groups in pmα-CD, effectively removing the intramolecular hydrogen bonding and increasing the hydration number and solubility of pmα-CD relative to α-CD [49]. β-CD also shows a higher probability of intramolecular hydrogen bonding than α- and γ-CD [8], consistent with the fact that β-CD is the least flexible. The effect of CDs on the dynamics and behaviour of the surrounding water molecules is another significant factor to investigate [43, 50, 52, 54-57]. For example, β-CD imposes a greater structuring on the local water molecules than α- and γ-CD [52], a phenomenon which can be linked to their flexibility trend, that is, the least flexible β-CD is not very easily incorporated into the water structure (i.e. not easily solvated) and therefore structures the water the most [52]. A greater structuring of water implies that the entropy of solution for β-CD would be the least favourable – this is consistent with previous entropy-enthalpy studies [50, 58], where the low solubility of β-CD has been linked to its unfavourable entropy of

solution. Factors contributing to the relative conformational flexibilities and solubilities of CDs are discussed in more depth in Chapter 6.

1.3.3 General Applications

Due to their complexing abilities, CDs are used in a variety of industrial products such as pharmaceuticals, foods, cosmetics and various chemical products [3]. A summary of some of the applications of CDs is given below (for a more in depth discussion of CD applications, I refer the reader to references [2-7]).

Solubility

CDs can alter the solubility of the complexed guest molecule, either by increasing or decreasing its solubility in a particular solvent [3]. For example:

- CDs and their derivatives are particularly important in the pharmaceutical industry, where they can be used to increase the aqueous solubility of drugs [6]. This can be explained by considering the structure of CDs (as discussed previously in Section 1.3.1), where the interior hydrophobic cavity of the CD binds the drug molecule, while the exterior hydrophilic surface of the CD makes the complex water-soluble as a whole. CDs are also used to increase the stability and bioavailability of drugs [4, 6].

- CDs can be used to increase the solubility of dyes in water, ensuring that more dye binds to the fabric as well as a reduction in the amount of dye left over in the waste-water [3].

- CDs can also be used to decrease the solubility of drugs, such as nonsteroidal anti-inflammatory drugs (NSAIDs). For example, the NSAID flufenamic acid is soluble in water and if administered in an uncomplexed form, would cause a sudden increase in the blood concentration and bring on unpleasant gastrointestinal side effects. To solve this problem, the aqueous solubility of flufenamic acid can be decreased by complexing it with a triacetyl β-CD derivative which is insoluble in water, ensuring slow release of the drug into the blood stream [3].

- CDs are also used to control the viscosity of water-based paints during the manufacturing process. Thickeners are added to the paint to increase its viscosity, but this makes mixing more difficult during the manufacturing process, thus the thickeners are initially complexed to CDs to reduce their solubility in the paint and therefore decrease the viscosity of the paint. Later, additional components are added to the paint to displace the thickener from the CD cavity and bring about the required viscosity to the paint [3].

- Hespiridin is a substance found in the juice of certain oranges and causes the juice of canned orange slices to turn cloudy. To remove the cloudiness, β-CD is added to the juice, which complexes and solvates the hesperidin. The undesirable taste of the hespiridin is also masked by complexation with β-CD [3].

Reactivity

CDs can be used to modify the reactivity of compounds. In the context of organic reactions, there are two ways in which the reactivity can be changed:

1. CDs can be used to *catalyse* the reaction, in which case they are often referred to as "artificial enzymes". The general case involves the CD and reactant first forming a CD-reactant intermediate through covalent bonding, after which the reaction proceeds to the final product [5]. The catalytic effects of CDs and their derivatives has been studied widely in literature [7]. For example, the Diels-Alder reaction between cyclopentadiene and a small dienophile such as acrylonitrile can be accelerated by β-CD, where both the reactants bind in the β-CD cavity for the addition reaction to take place [7]. The regioselectivity of a Diels-Alder reaction can also be promoted by the reaction taking place within the cavity of a CD [7].

2. CDs can also be used to *mediate* organic reactions. In this case, there is no covalent bonding between the CD and reactant, and instead the hydrophobic cavity of the CD provides a new environment for the reaction to take place, where aspects of the reactivity (e.g. the rate or selectivity of the reaction) can be

changed [5]. The CD cavity can therefore also be referred to as an "*extra reaction field*" in this context [5].

Stability

CDs can also be used to stabilise the guest compound against heat, light and oxidation effects [3]. For example:

- Nicardipine is a compound that decomposes when exposed to light, but its rate of tessellation can be reduced by complexing it with CDs [3]. The rate of photodegradation decreases by a factor of 10 when nicardipine is complexed with methylated β-CD, 6.5 with α-CD and 5 with γ-CD [3].

- Unsaturated fats are a component of several oils such as vegetable, grapeseed and olive oils. However, they contain unsaturated fatty acids that are easily oxidised, causing unpleasant tastes and smells – complexing them with CDs protects them against oxidation [3].

- Peroxy acids are used in disinfectants and bleaching solutions, which are commonly applied in the home, industry, and to clean medical equipment. Complexing the peroxy acids with CDs improves their thermal stability and storage, and also reduces their odour [3].

Volatility

The volatility of compounds can also be reduced by complexation with CDs, as the interactions between the CD and guest molecule increase the energy barrier the guest needs to overcome in order to become volatile [3]. For example:

- The volatility of menthol can be decreased by complexing it with β-CD. Very little menthol is lost from the complex, even after drying it at 100 °C, whereas the same drying conditions would completely volatilize free (uncomplexed) menthol [3].

- CDs are also used to complex perfumes in order to prolong their release into the air [3].

Spectral properties

The spectral properties of a compound can also be modified by complexation with CDs [2, 3]. For example:

- The intensity of fluorescence or light absorption of some compounds can be either decreased or increased when the compound is complexed with a CD, which can be very useful in analytical chemistry applications [3]. An example of a specific application in photography is where the intensity of fluorescence of an optical brightener was increased by complexation with a methylated CD, which caused the white areas of the photograph to become whiter [3].

- The chemical shifts in the NMR spectra of the compound can be changed by complexation with CDs [2]. The optical activity of a compound can also be changed, i.e. the CD cavity is chiral, and when achiral compounds are complexed with the CD, the compounds become optically active [2].

1.4 Objectives

The main objective of this thesis is to use the triangular tessellation model [9] to quantitatively describe the macrocyclic ring puckering motions of α-, β- and γ-CD in water from dynamic computer simulations. The puckering of the CDs is studied through a *coarse grained analysis of full atomistic simulations*, where the *macrocyclic* conformational behaviour is considered rather than the atomistic detail of the simulations. APC based on the TT model are used in this work since (i) they can be applied relatively easily to the *macrocyclic* CD rings through a suitable coarse graining of the CDs, and (ii) the APC are all the same type of quantity (i.e. angular) and the conformational motions across the *homogeneous* CD rings are measured over a long simulation time, permitting *average* time correlations to be calculated for the APC of each CD – these correlations will then be compared to obtain information about the relative CD flexibilities. The flexibilities will then be compared to the CD solubility trend, to help explain the anomalous aqueous solubility of β-CD, as compared to α- and γ-CD. Conformational changes are important in helping to explain physical properties of ring systems (such as solubility) – for this reason, the *ring puckering* of CDs is being analysed here, to help explain their solubility trend. β-CD is industrially very important, thus exploring molecular reasons for its low solubility (its inhibiting property) may help, in future, to design more effective cyclodextrin-based industrial products.

In Chapter 2, the theory of computer simulations will be discussed, with emphasis on molecular dynamics (MD) simulations.

Chapter 3 describes analytical techniques that are used to study data generated from computer simulations. In particular, the statistical mechanics technique of *time correlation functions* will be discussed. Time correlation functions are used together with APC in this thesis to monitor the dynamic behaviour and rate of decay of the CD puckering motion.

Chapter 4 presents a background to qualitative and quantitative methods of analysing ring pucker. The focus is on quantitative puckering models based on (i) perpendicular displacements of the ring atoms from a mean plane, (ii) the intracyclic torsion angles, and (iii) the triangular tessellation of the ring.

In Chapter 5, the TT definition for 6-membered rings will be extended to 7- and 8-membered rings, including derivations for the APC of 7- and 8-membered rings. A

systematic procedure is also developed to generate the APC of all the states of cycloheptane and cyclooctane.

In Chapter 6, the relative conformational flexibilities of α-, β- and γ-CD in water will be investigated through MD simulations. The macrocyclic ring puckering of the CDs is measured from the MD trajectories using the TT definition, and then analysed using probability distribution plots and time correlation functions. The relationship between CD flexibility and aqueous solubility is then investigated, using the rate of decay of the puckering motion (as obtained from the time correlation functions) as a measure of the CD flexibilities. In addition, the macrocyclic "breathing" motion of the CD rings (planar deformation motion) is defined and analysed from the trajectories to supplement the puckering analysis.

References

1. Sega, M.; Autieri, E.; Pederiva, F. *The Journal of Chemical Physics* **2009**, *130*, 225102.
2. Szejtli, J. *Chemical Reviews* **1998**, *98*, 1743.
3. Hedges, A. R. *Chemical Reviews* **1998**, *98*, 2035.
4. Uekama, K.; Hirayama, F.; Irie, T. *Chemical Reviews* **1998**, *98*, 2045.
5. Takahashi, K. *Chemical Reviews* **1998**, *98*, 2013.
6. Loftsson, T.; Brewster, M. E. *Journal of Pharmaceutical Sciences* **1996**, *85*, 1017.
7. Breslow, R.; Dong, S. D. *Chemical Reviews* **1998**, *98*, 1997.
8. Naidoo, K. J.; Gamieldien, M. R.; Chen, J. Y.-J.; Widmalm, G. r.; Maliniak, A. *The Journal of Physical Chemistry B* **2008**, *112*, 15151.
9. Hill, A. D.; Reilly, P. J. *Journal of Chemical Information and Modeling* **2007**, *47*, 1031.
10. Stoddart, J. F. *Stereochemistry of Carbohydrates*; John Wiley & Sons, Inc.: New York, 1971.
11. Bérces, A.; Whitfield, D. M.; Nukada, T. *Tetrahedron* **2001**, *57*, 477.
12. Saenger, W. *The principles of Nucleic acid structure*; Springer: Heidelberg, 1988 (Second Corrected Printing).
13. Marszalek, P. E.; Oberhauser, A. F.; Pang, Y.-P.; Fernandez, J. M. *Nature* **1998**, *396*, 661.
14. Ionescu, A. R.; Bérces, A.; Zgierski, M. Z.; Whitfield, D. M.; Nukada, T. *The Journal of Physical Chemistry A* **2005**, *109*, 8096.
15. Lindoy, L. F. *The Chemistry of Macrocyclic Ligand Complexes*; Cambridge University Press: Cambridge, 1989.
16. Seiler, P.; Dobler, M.; Dunitz, J. D. *Acta Crystallographica Section B* **1974**, *30*, 2744.
17. Dobler, M.; Dunitz, J. D.; Seiler, P. *Acta Crystallographica Section B* **1974**, *30*, 2741.
18. Sinnott, M. L. *Chemical Reviews* **1990**, *90*, 1171.
19. Vasella, A.; Davies, G. J.; Böhm, M. *Current Opinion in Chemical Biology* **2002**, *6*, 619.
20. Davies, G.; Henrissat, B. *Structure* **1995**, *3*, 853.
21. Barnett, C. B.; Naidoo, K. J. *Molecular Physics* **2009**, *107*, 1243.

22. Biarnés, X.; Ardèvol, A.; Planas, A.; Rovira, C.; Laio, A.; Parrinello, M. *Journal of the American Chemical Society* **2007**, *129*, 10686.
23. Davies, G. J.; Ducros, V. M.-A.; Varrot, A.; Zechel, D. L. *Biochemical Society Transactions* **2003**, *31*, 523.
24. Fushinobu, S.; Mertz, B.; Hill, A. D.; Hidaka, M.; Kitaoka, M.; Reilly, P. J. *Carbohydrate Research* **2008**, *343*, 1023.
25. Schramm, V. L. *Annual Review of Biochemistry* **1998**, *67*, 693.
26. Gloster, T. M.; Davies, G. J. *Organic & Biomolecular Chemistry* **2010**, *8*, 305.
27. Brayer, G. D.; Sidhu, G.; Maurus, R.; Rydberg, E. H.; Braun, C.; Wang, Y.; Nguyen, N. T.; Overall, C. M.; Withers, S. G. *Biochemistry* **2000**, *39*, 4778.
28. Bocian, D. F.; Pickett, H. M.; Rounds, T. C.; Strauss, H. L. *Journal of the American Chemical Society* **1975**, *97*, 687.
29. Cremer, D.; Pople, J. A. *Journal of the American Chemical Society* **1975**, *97*, 1354.
30. Kilpatrick, J. E.; Pitzer, K. S.; Spitzer, R. *Journal of the American Chemical Society* **1947**, *69*, 2483.
31. Pickett, H. M.; Strauss, H. L. *The Journal of Chemical Physics* **1971**, *55*, 324.
32. Altona, C.; Geise, H. J.; Romers, C. *Tetrahedron* **1968**, *24*, 13.
33. Geise, H. J.; Altona, C.; Romers, C. *Tetrahedron Letters* **1967**, *8*, 1383.
34. Diez, E.; Esteban, A. L.; Bermejo, F. J.; Rico, M. *The Journal of Physical Chemistry* **1980**, *84*, 3191.
35. Haasnoot, C. A. G. *Journal of the American Chemical Society* **1992**, *114*, 882.
36. Zefirov, N. S.; Palyulin, V. A. *Dokl. Akad. Nauk SSSR.* **1980**, *252*, 111.
37. Zefirov, N. S.; Palyulin, V. A.; Dashevskaya, E. E. *Journal of Physical Organic Chemistry* **1990**, *3*, 147.
38. Strauss, H. L.; Pickett, H. M. *Journal of the American Chemical Society* **1970**, *92*, 7281.
39. Joshi, N. V.; Rao, V. S. R. *Biopolymers* **1979**, *18*, 2993.
40. Saenger, W.; Jacob, J.; Gessler, K.; Steiner, T.; Hoffmann, D.; Sanbe, H.; Koizumi, K.; Smith, S. M.; Takaha, T. *Chemical Reviews* **1998**, *98*, 1787.
41. Lipkowitz, K. B. *Chemical Reviews* **1998**, *98*, 1829.
42. Dodziuk, H. *Journal of Molecular Structure* **2002**, *614*, 33.
43. Pereira, C. S.; de Moura, A. F.; Freitas, L. C. G.; Lins, R. D. *Journal of the Brazilian Chemical Society* **2007**, *18*, 951.

44. Kozár, T.; Venanzi, C. A. *Journal of Molecular Structure: THEOCHEM* **1997**, *395–396*, 451.
45. Thaning, J.; Stevensson, B.; Östervall, J.; Naidoo, K. J.; Widmalm, G. r.; Maliniak, A. *The Journal of Physical Chemistry B* **2008**, *112*, 8434.
46. Schneider, H.-J.; Hacket, F.; Rüdiger, V.; Ikeda, H. *Chemical Reviews* **1998**, *98*, 1755.
47. Harata, K. *Chemical Reviews* **1998**, *98*, 1803.
48. Rekharsky, M. V.; Inoue, Y. *Chemical Reviews* **1998**, *98*, 1875.
49. Shikata, T.; Takahashi, R.; Satokawa, Y. *The Journal of Physical Chemistry B* **2007**, *111*, 12239.
50. Linert, W.; Margl, P.; Renz, F. *Chemical Physics* **1992**, *161*, 327.
51. Bernatowicz, P.; Ruszczyńska-Bartnik, K.; Ejchart, A.; Dodziuk, H.; Kaczorowska, E.; Ueda, H. *The Journal of Physical Chemistry B* **2010**, *114*, 59.
52. Naidoo, K. J.; Chen, J. Y.-J.; Jansson, J. L. M.; Widmalm, G.; Maliniak, A. *The Journal of Physical Chemistry B* **2004**, *108*, 4236.
53. Perez-Miron, J.; Jaime, C.; Ivanov, P. M. *Chirality* **2008**, *20*, 1127.
54. Raffaini, G.; Ganazzoli, F. *Chemical Physics* **2007**, *333*, 128.
55. Starikov, E. B.; Bräsicke, K.; Knapp, E. W.; Saenger, W. *Chemical Physics Letters* **2001**, *336*, 504.
56. Heine, T.; Dos Santos, H. F.; Patchkovskii, S.; Duarte, H. A. *The Journal of Physical Chemistry A* **2007**, *111*, 5648.
57. Rodriguez, J.; Hernán Rico, D.; Domenianni, L.; Laria, D. *The Journal of Physical Chemistry B* **2008**, *112*, 7522.
58. Jozwiakowski, M. J.; Connors, K. A. *Carbohydrate Research* **1985**, *143*, 51.

Chapter 2

Molecular Modelling

Computer simulations form an important part of chemistry, with the aid of principles from physics to model chemical systems and simulate their behaviour. In dynamic simulations, the interactions between particles are calculated using a physical model to generate detailed information about the system on a *microscopic* scale (i.e. the positions and velocities of all the individual particles). This information can then be related to *macroscopic* properties of the system, such as transport coefficients, internal energy, and heat capacities [1], using techniques of statistical mechanics. It can be very difficult to analyse microscopic motions experimentally, especially for large systems such as solvated macromolecules – this is where computer simulations can become very useful, giving insight into the dynamics and workings of enzymes, heterogeneous catalysts, and other biologically important systems [1, 2].

2.1 Molecular Modelling Approaches

There are two basic classes of molecular modelling simulations, namely those using ***quantum mechanical*** (QM) methods, which include *ab initio* and semi-empirical simulations, and those using ***force field methods*** (also called ***Molecular Mechanics*** or ***MM methods***) [3]. The dynamics of chemical systems can also be simulated, for example, molecular dynamics (MD) simulations can be run on MM and QM/MM systems. Different types of MD packages have been developed, including CHARMM [4], AMBER [5], GROMOS [6], and DL_POLY [7]. An overview of QM and force field methods will be given in the following sections. In this thesis, the CHARMM force field [4] is used to conduct MD simulations.

2.2 Quantum Mechanics

Quantum mechanics provides a complete description of chemical systems on the atomic and molecular scale, and is very useful to investigate the breaking and forming of bonds in chemical reactions. Moving particles can be represented with a wavefunction, which provides a complete description of its state and motion. From the wavefunction, different properties of the system can be calculated, such as its energy, momentum, and

spatial probability distribution [3, 8]. One of the central equations of quantum mechanics is the Schrödinger equation, which relates the total energy E of an atomic or molecular system to its wavefunction Ψ – the time-independent form of the Schrödinger equation is given in Equation **2.1** [9]:

$$H_{\text{tot}}\Psi_{\text{tot}}(\{r_i, R_i\}) = E_{\text{tot}}\Psi_{\text{tot}}(\{r_i, R_i\}) \quad (\mathbf{2.1})$$

where H_{tot} is the molecular Hamiltonian (the operator), Ψ_{tot} is the molecular wavefunction (the eigenfunction), E_{tot} is the total kinetic and potential energy of the molecule (the eigenvalue), and $\{r_i\}$ and $\{R_i\}$ are the electronic and nuclear positional coordinates, respectively. H_{tot} can also be broken down as in Equation **2.2** [9]:

$$H_{tot} = T_{elec}(\{r_i\}) + T_{nucl}(\{R_i\}) + V_{elec-elec}(\{r_i\}) + V_{nucl-nucl}(\{R_i\}) + V_{elec-nucl}(\{r_i, R_i\}) \quad (\mathbf{2.2})$$

where T_{elec} and T_{nucl} are the electronic and nuclear kinetic energies, respectively, and $V_{elec\text{-}elec}$, $V_{nucl\text{-}nucl}$, and $V_{elec\text{-}nucl}$ are the electrostatic potential energies of electron-electron repulsion, nucleus-nucleus repulsion and electron-nucleus attraction.

The Schrödinger equation can be solved for values of Ψ and E. However, for large systems such as biological macromolecules, solving the Schrödinger equation is a complicated and computationally demanding task. To simplify this problem, the Born-Oppenheimer (BO) approximation is used [9], which states that since nuclei are much heavier and move slower than electrons, the motions of electrons and nuclei can be treated separately. Under this approximation, Ψ_{tot} can be written approximately as a product of electronic and nuclear wavefunctions, as given in Equation **2.3** [9].

$$\Psi_{\text{tot}} = \Psi_{elec}\Psi_{nucl} \quad (\mathbf{2.3})$$

Under the BO approximation, the Schrödinger equation and molecular Hamiltonian can also be separated into electronic and nuclear parts. The electronic parts are considered first – these are given by the electronic Hamiltonian and electronic Schrödinger equation in Equations **2.4** and **2.5**, respectively [9]. Here, the electrons are moving

while the nuclei are considered to be stationary, i.e. the T_{nucl} term of Equation **2.2** is neglected in the expression for H_{elec}. $V_{nucl\text{-}nucl}$ is a constant and is also neglected in Equation **2.4** as adding a constant to H_{elec} only affects the eigenvalues, and not the eigenfunctions [10].

$$H_{elec} = T_{elec}(\{r_i\}) + V_{elec-elec}(\{r_i\}) + V_{elec-nucl}(\{r_i, R_i\}) \quad (\mathbf{2.4})$$

$$H_{elec}\Psi_{elec}(\{r_i, R_i\}) = E_{elec}(\{R_i\})\Psi_{elec}(\{r_i, R_i\}) \quad (\mathbf{2.5})$$

The electronic Schrödinger equation is solved first to give a set of electronic wavefunctions Ψ_{elec} and a corresponding set of electronic energies E_{elec}. Ψ_{elec} depends explicitly on $\{r_i\}$ and *parametrically* on the nuclear coordinates $\{R_i\}$, that is, Ψ_{elec} depends on the nuclei positions but not on their velocities (as the nuclei are considered to be fixed). The nuclei positions can then be changed by an infinitesimal amount and the electronic Schrödinger equation can be solved again to generate a new set of Ψ_{elec} – in this way, sets of Ψ_{elec} and corresponding E_{elec} values can be generated for several different nuclear configurations of the system. The total energy of each nuclear configuration (i.e. the sum of E_{elec} and $V_{nucl\text{-}nucl}$) can then be plotted against the configuration to generate a Potential Energy Surface (PES), from which important chemical information can be obtained, such as the dynamic behaviour of chemical reactions. The BO approximation is usually good – for the hydrogen molecule, the errors in calculating the energy is in the order of 10^{-4}, and gets better for larger systems with heavier nuclei [8, 9].

The next step is to consider the motion of the nuclei and use the PES to solve the nuclear Schrödinger equation. The nuclei move in a total potential V_{tot}, given in Equation **2.6**. V_{tot} is made of two parts, namely $V_{nucl\text{-}nucl}$ together with an *average* potential induced by the much faster-moving electrons – the potential induced by the electrons is just the values of E_{elec} averaged over time, as shown in Equation **2.7**. Thus the resulting nuclear Hamiltonian H_{nucl} includes V_{tot} together with the nuclear motion T_{nucl} (Equation **2.8**), and the nuclear Schrödinger equation (Equation **2.9**) can then be solved for the nuclear wavefunction Ψ_{nucl} and the total molecular energy E_{tot} [9, 10].

$$V_{tot}(\{R_i\}) = \langle E_{elec}(\{R_i\})\rangle + V_{nucl-nucl}(\{R_i\}) \quad (\mathbf{2.6})$$

$$\langle E_{elec}(\{R_i\})\rangle = \langle T_{elec} + V_{elec-elec} + V_{elec-nucl}(\{R_i\})\rangle \quad (\mathbf{2.7})$$

$$H_{nucl}(\{R_i\}) = T_{nucl}(\{R_i\}) + V_{tot}(\{R_i\}) \quad (\mathbf{2.8})$$

$$H_{nucl}\Psi_{nucl}(\{R_i\}) = E_{\text{tot}}(\{R_i\})\Psi_{nucl}(\{R_i\}) \quad (\mathbf{2.9})$$

The functions Ψ_{nucl} describe all the nuclear motions (translational, rotational and vibrational) of the molecule, from which macroscopic properties of the system can be calculated [8, 11].

2.3 Force Field Methods

The Schrödinger equation is very complex to solve for very large systems [9], thus QM methods are computationally intensive. To simplify the problem, force field methods are used to model larger systems (such as solvated macromolecules) – these methods use simpler models based on the laws of *classical mechanics*, thereby decreasing the time needed for the simulations. However, there are some limitations in that classical mechanics models are not suited for simulating chemical reactions, processes taking place at low temperatures (0-10 K), or simulating the motion of hydrogen atoms [12].

2.3.1 Potential Energy Function

In force field models, the motion of the electrons are ignored and the potential energy of the system is written as a function of the nuclear coordinates only – this can be done due to the BO approximation explained earlier [3]. The potential energy function, or *force field*, is defined by two basic parts, namely, the functional form of the energy terms and the numerical parameters used in these terms (the parameters can be obtained experimentally or by using *ab initio* methods) [3, 12].

In MM methods, the system is described using the classical "ball and spring" model [9], where the atoms are considered to be hard spheres connected to each other by springs. The potential energies associated with the various motions of the system (such as bond-stretching, rotation about single bonds, etc.) are included in the potential energy

function, which is the sum total of all the energy contributions, indicating the relative stability of each configuration. The potential energy function, such as the one used in CHARMM [4], has the following form:

$$E = E_{bonds} + E_{angles} + E_{dihedrals} + E_{improper-dihedrals} + E_{Urey-Bradley} + E_{VdW} + E_{elec} \quad (\mathbf{2.10})$$

The terms in Equation **2.10** can be divided into two categories, namely *intramolecular* and *intermolecular* (or *non-bonded*) terms [13]. The intramolecular terms arise from the motions within the molecular system itself, while the intermolecular terms are due to interactions between the molecule and other particles, such as solvent particles. The terms are discussed in detail below, and most are illustrated in **Figure 2.2**.

***Intramolecular* energy terms**

- **Bond potential**

The bond potential gives the energy associated with the stretching of bonds within the molecule, and is given by Equation **2.11** [13]:

$$E_{bonds} = \sum_{\text{bonds}} K_b(b - b_0)^2 \quad (\mathbf{2.11})$$

b_0 is the equilibrium (ideal) bond length, and K_b is the force constant which indicates the strength of the bond – the larger the value of K_b, the more energy it takes to stretch the bond from its equilibrium length. K_b is different for different bond types. Values for K_b can be determined from vibrational data (such as IR spectroscopy), and values for b_0 can be obtained from crystallographic or microwave data [14].

- **Bond angle potential**

The bond angle potential measures the energy associated with the bending of the angles defined by three bonded atoms, and is given as follows [13]:

$$E_{angles} = \sum_{\text{angles}} K_\theta(\theta - \theta_0)^2 \quad (\mathbf{2.12})$$

θ_0 is the equilibrium bond angle and K_θ is the force constant for the angle-bending motion. It takes less energy to bend bond angles than to stretch bonds, thus the K_θ force constants tend to be smaller than K_b values [3].

The two terms above are referred to as 'hard' degrees of freedom, because it usually takes a lot of energy to cause significant deviations of the bond lengths and angles from the equilibrium values [3]. At normal temperatures and when there are no chemical reactions taking place, the deviations are small enough to express them using harmonic potential (Hooke's law) functional forms [4], as shown in Equations **2.11** and **2.12**.

- **Dihedral angle potential**

When four atoms are bonded in sequence, the proper dihedral angle is the angle of rotation about the axis running through the bond between the middle pair of atoms. The dihedral angle potential measures the energy associated with this rotation, and is expressed as follows [13]:

$$E_{dihedrals} = \sum_{\text{dihedrals}} K_\phi(1 + \cos(n\phi - \delta)) \qquad \textbf{(2.13)}$$

where ϕ and K_ϕ represent the dihedral angle and force constant, respectively. n and δ are the dihedral multiplicity and phase angle, respectively.

- **Improper torsions potential**

The improper dihedral (torsion) is defined by 4 atoms *i*, *j*, *k*, and *l* arranged such that one of the atoms *j* is directly bonded to the other three *i*, *k* and *l*, so that *j* forms the central atom in the arrangement – this is depicted in **Figure 2.1**.

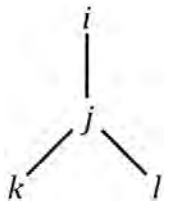


Figure 2.1: Schematic of the atom arrangements for the improper dihedral angle.

The improper dihedral is the angle formed between planes *jkl* and *ikl*, thus it measures the deviation of the atom arrangement from the planar configuration [4]. The potential

associated with changes in improper dihedral angles is given in Equation **2.14** below [4], where ω, ω_0 and K_ω are the improper dihedral angle, equilibrium improper dihedral, and force constant, respectively.

$$E_{improper-dihedrals} = \sum_{\text{improper}-\text{dihedrals}} K_\omega(\omega - \omega_0)^2 \qquad \textbf{(2.14)}$$

- **Urey-Bradley potential**

When three atoms are bonded covalently, changes in the distance between the first and third atoms (or 1,3 distance) can be modelled by the Urey-Bradley potential, which is a harmonic potential similar in form to the bond and bond angle potentials. The Urey-Bradley potential is given in Equation **2.15** [13].

$$E_{Urey-Bradley} = \sum_{\text{Urey}-\text{Bradley}} K_{UB}(r_{1,3} - r_{1,3;0})^2 \qquad \textbf{(2.15)}$$

K_{UB} is the equilibrium constant, $r_{1,3}$ is the 1,3 distance and $r_{1,3;0}$ is the equilibrium 1,3 distance.

Intermolecular energy terms

The intermolecular terms model the interactions between all non-bonded pairs of atoms, which are either part of different molecules or part of the same molecule but separated by three or more bonds [3] – in the latter case, the atoms are not strictly *intermolecular* but still have non-bonded interactions. The terms involve van der Waals and electrostatic interactions, as outlined below.

- **van der Waals (vdW) interactions**

The van der Waals interactions are usually modelled using the Lennard-Jones 12-6 potential, given Equation **2.16** [13].

$$E_{van\ der\ Waals} = \sum_{\text{non}-\text{bonded pairs}} \varepsilon_{ij}\left[\left(\frac{R_{\min,ij}}{r_{ij}}\right)^{12} - 2\left(\frac{R_{\min,ij}}{r_{ij}}\right)^{6}\right] \qquad \textbf{(2.16)}$$

The interacting pair of atoms are designated *i* and *j*, r_{ij} is the distance between the atoms, ε_{ij} is the well-depth, and $R_{\min,ij}$ is the distance between the atoms at which the potential energy is a minimum. An alternate and equivalent form of the Lennard-Jones

potential is given in Equation **2.17** (adapted from reference [3]):

$$E_{van\ der\ Waals} = \sum_{\text{non-bonded pairs}} \left(\frac{A_{ij}}{{r_{ij}}^{12}} - \frac{C_{ij}}{{r_{ij}}^{6}} \right) \qquad (\mathbf{2.17})$$

where $A = (\varepsilon_{ij})(R_{\min,ij})^{12}$ and $C = 2(\varepsilon_{ij})(R_{\min,ij})^{6}$.

- **Electrostatic interactions**

Electrostatic interactions occur between charged species (ions or atoms with partial charges) and are usually modelled as a Coulomb potential, as in Equation **2.18** [13].

$$E_{electrostatic} = \sum_{\text{non-bonded pairs}} \frac{q_i q_j}{4\pi D r_{ij}} \qquad (\mathbf{2.18})$$

q_i and q_j are the charges or partial atomic charges on atoms i and j, r_{ij} is the distance between the atoms, and D is the dielectric constant of the medium in which the simulation takes place. **Figure 2.2** depicts the main interaction terms in the potential energy function.

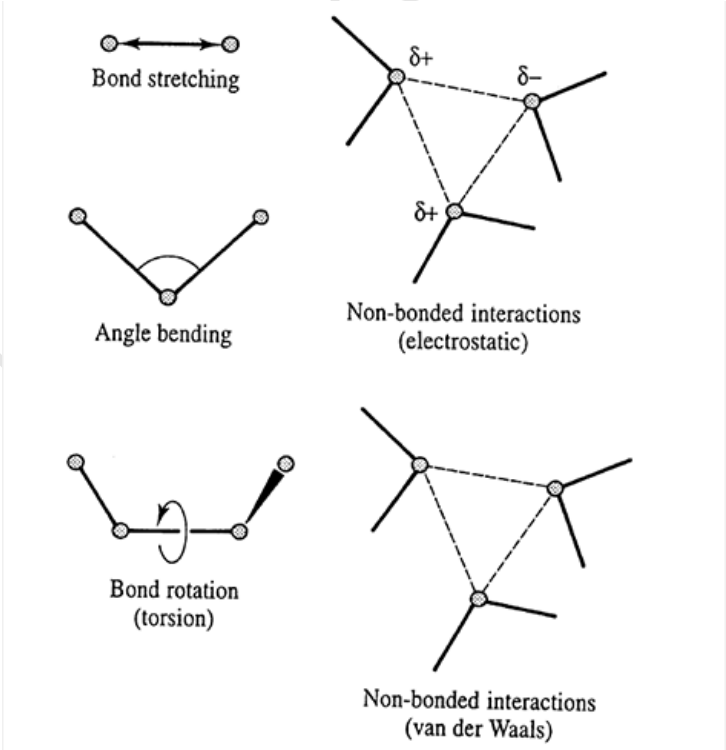


Figure 2.2: Representation of the main interactions contributing to a force field: bond-stretching, angle-bending, torsional bond rotations, and non-bonded interactions (taken from reference [3]).

2.3.2 Parameterisation of Force Field

There are many different force fields in literature, which differ in the types and functional form of the energy terms, the type of system being studied (carbohydrates, proteins, etc.), and the nature of the simulation (whether it is carried out in the gas phase or in a solvent) [12]. Once a force field function is developed, it then needs to be parameterised to accurately model the system under study – parameters such as equilibrium constants, equilibrium bond-angles and distances, Lennard-Jones parameters, and electrostatic parameters need to be evaluated. There are two main ways to parameterise a force field:

1. The parameters can be fit to experimental data such as X-ray crystal structures, IR data, density measurements, enthalpies of vaporisation, NMR data, and free energies of solvation [12].
2. The parameters can also be fit to data obtained from *ab initio* QM calculations – this is useful when experimental data is hard to get for the types of molecules being parameterised [3, 12].

2.4 Molecular Dynamics

MD simulations use a parameterised force field together with classical methods to generate a trajectory of all the atomic motions in the system over time [12]. Using statistical mechanics techniques, various time-dependent properties can be extracted from MD trajectories, giving insight into the *dynamical* behaviour of a system.

2.4.1 Integrating the Equations of Motion

To generate the trajectories, Newton's equations of motion (Equations **2.19** and **2.20**) are integrated simultaneously for all atoms in the system [12].

$$\frac{\mathrm{d}^2\boldsymbol{r}_i(t)}{\mathrm{d}t^2} = \frac{\boldsymbol{F}_i}{m_i} \tag{2.19}$$

$$\boldsymbol{F}_i = -\frac{\partial V(\boldsymbol{r}_i, \ldots, \boldsymbol{r}_N)}{\partial \boldsymbol{r}_i} \tag{2.20}$$

Equation **2.19** is Newton's second law, where $\boldsymbol{F}_i$ is the total force on atom i (due to interactions with the other atoms in the system), $\boldsymbol{r}_i$ and m_i are the positional coordinates and mass of atom i respectively, t represents time, and $\mathrm{d}^2\boldsymbol{r}_i(t)/\mathrm{d}t^2$ is the acceleration of atom i (abbreviated $\boldsymbol{a}_i$). $V(\boldsymbol{r}_i\ ,\ \ldots\ ,\ \boldsymbol{r}_N)$ is the force field potential energy of a given configuration of the system (where N is the total number of atoms), which is related to the force on atom i through Equation **2.20**.

The actual integration is carried out in small time steps Δt, which usually has a value between 1 and 10 fs [12]. The basic procedure for the integration can be outlined as follows [3]:

1. The potential V is differentiated to give $\boldsymbol{F}_i$ at a time t using Equation **2.20**.
2. $\boldsymbol{F}_i$ is then used to get the acceleration $\boldsymbol{a}_i$ using Equation **2.19**.
3. $\boldsymbol{a}_i$ is combined with the position and/or velocity of atom i to give the new position and/or velocity at time $t + \Delta t$. $\boldsymbol{F}_i$ (and therefore $\boldsymbol{a}_i$) are considered to be constant over the duration of the time step. How the new positions are actually obtained depends on the integration algorithm that is used.
4. The above steps are carried out for all the atoms i in the system, generating a configuration of the system. $\boldsymbol{F}_i$ and $\boldsymbol{a}_i$ are then calculated using the positions evaluated at time $t + \Delta t$, from which the new positions for time $t + 2\Delta t$ are obtained. This procedure is repeated for all the time steps in the simulation.

The sequential configurations of the system evaluated at each time step over the length of the simulation form the MD trajectory. Different integration algorithms can be used in Step 3 above. One of the more commonly used methods is the *Verlet algorithm* [15], where the positions at times t and $t - \Delta t$ (previous time step) are used together with the acceleration at time t to give the new position at time $t + \Delta t$. The Verlet algorithm is summarised in Equation **2.21** below [3].

$$\boldsymbol{r}(t + \Delta t) = 2\boldsymbol{r}(t) - \boldsymbol{r}(t - \Delta t) + \Delta t^2\boldsymbol{a}(t) \qquad (\mathbf{2.21})$$

One disadvantage of the Verlet algorithm is that adding the small term $\Delta t^2\boldsymbol{a}(t)$ to the larger term of $(2\boldsymbol{r}(t) - \boldsymbol{r}(t - \Delta t))$ causes a loss of precision in the calculation of the position at $t + \Delta t$ [3].

A development on the Verlet algorithm is the *Leap-Frog algorithm* [16], which has been used for the simulations in this thesis. The Leap-Frog algorithm is shown in Equations **2.22** and **2.23** below [1], where velocities $\boldsymbol{v}$ at different times are also included in the calculation of the new positions.

$$\boldsymbol{r}(t + \Delta t) = \boldsymbol{r}(t) + \Delta t \boldsymbol{v}\left(t + \tfrac{1}{2}\Delta t\right) \qquad (\mathbf{2.22})$$

$$\boldsymbol{v}\left(t + \tfrac{1}{2}\Delta t\right) = \boldsymbol{v}\left(t - \tfrac{1}{2}\Delta t\right) + \Delta t \boldsymbol{a}(t) \qquad (\mathbf{2.23})$$

First, Equation **2.23** is used to calculate $\boldsymbol{v}(t + ½\Delta t)$, which is then used in Equation **2.22** with $\boldsymbol{r}(t)$ to obtain the new position $\boldsymbol{r}(t + \Delta t)$. Repeating this process, the velocity at $t + ½\Delta t$ "leap-frogs" over the position at $t + \Delta t$ to give $\boldsymbol{v}(t + 1½\Delta t)$, then the position at $t + \Delta t$ "leap-frogs" over this velocity to give $\boldsymbol{r}(t + 2\Delta t)$. One advantage of the Leap-Frog algorithm over the Verlet method is that in the former, there are no very small terms (such as terms containing Δt^2) being added to larger terms, thus making the Leap-Frog algorithm more precise [1]. For MD simulations of liquids and solutions, the Leap-Frog algorithm is reported to be a very accurate and efficient method to use [12].

2.4.2 Carbohydrate Force Fields

Carbohydrates play an important role in many biological systems, and are generally very flexible molecules, especially in solution. To model them accurately can be challenging, as the model should simulate their conformational motions and flexibility in a manner that is in agreement with experimental data [17]. Several force fields have been parameterised for carbohydrates – some of those developed for CHARMM are outlined below.

The first CHARMM force field parameterised for carbohydrates is called the HGFB force field [18], where the letters 'HGFB' represent the authors' initials. This force field was developed for simple carbohydrates or monosaccharides in solution, with particular emphasis to reproduce experimental vibrational and X-ray data [19]. As an example, the model was tested on α-D-glucose and was found to reproduce its vibrational spectrum [18]. Some limitations of the HGFB force field include unrealistic modelling of ring

flexibility, as well as giving inaccurate conformational distributions of the primary alcohol (hydroxymethyl) group in pyranoses [17].

Improving upon the HGFB force field, Palma *et al.* [20] developed the PHLB force field and parameter set, where they replaced the general terms for dihedral angles in the HGFB force field with specific dihedral angle terms. The PHLB terms were then parameterised to reproduce experimental vibrational frequencies and to be consistent with *ab initio* energy maps for the dihedral angle rotations of relevant small molecules like ethylene glycol [21]. Although the PHLB force field predicts the order of the energies for the different primary alcohol conformations correctly, it does not accurately model the frequency of transitions between these conformers [17].

Then in 2002, Kuttel *et al.* [17] modified the PHLB force field to correct the rotational frequency of the primary alcohol group, and came up with the Carbohydrate Solution Force Field (CSFF). They tested the CSFF model by calculating potential of mean force (PMF) energy profiles for the primary alcohol group of the monosaccharides *β*-D-glucose and *β*-D-galactose in solution, which showed that both the primary alcohol rotational frequencies and equilibrium conformational distributions were in agreement with experimental data. In addition, the calculated vibrational frequencies for *α*-D-glucose in vacuum compared well with experimental frequencies, further validating the force field [17].

Other useful carbohydrate force fields have been developed in recent years that incorporate parameters for carbohydrates [22-25]. However, in this thesis, the CSFF force field will be used, the reasons for which will be explained in Chapter 6.

2.4.3 Steps in an MD Simulation

The steps involved in generating an MD trajectory, as carried out in CHARMM, are outlined below:

- *Minimisation*

  First, a starting structure of the system is chosen – this can be obtained from X-ray crystal data, or for simpler molecules, the structure can be constructed in CHARMM

or other structure building applications. The coordinates are then adjusted to minimise the potential energy of the system and to remove large forces (such as large repulsive forces from van der Waals interactions) which can cause unrealistic distortions later in the dynamics runs. In CHARMM, the minimisation is carried out using an iterative algorithm, such as the methods of steepest descents or conjugate gradients, among others [4].

- *Solvation*

  In the case of a solvated system, the molecule is usually minimised separately (in vacuum) and then solvated by specifying the number of solvent molecules and reading in their coordinates. The solvated system is then minimised again.

- *Heating*

  The minimised (solvated) system is then heated to the temperature chosen for the simulations. The heating is usually done gradually in small increments, as heating too quickly may cause the simulation to crash. The system is first set at a low temperature by assigning appropriate initial velocities to all the atoms – for example, this can be done according to a Gaussian probability distribution of velocities for that temperature. Molecular dynamics (using the specified integration algorithm) are run on the system for a short time, then the temperature is increased by a small increment by reassigning or scaling the velocities to match a higher temperature. This process is repeated until the final temperature is reached [4].

- *Equilibration*

  After heating, MD is usually run for a short time to equilibrate the system. Equilibration ensures that the temperature and fluctuations of the system [4] are stabilised before running dynamics.

- *Dynamics*

  After equilibration, MD is run for a longer period to generate the actual dynamical time trajectory of the system. Subsequent analysis of the trajectories can then be done, through statistical mechanics techniques for example.

2.4.4 Water Models

Simulations of solvated systems also require a suitable model to describe the structuring and interactions of the solvent molecules. An important solvent is water, for which different types of models have been developed. In simple models, the water molecules are maintained in a rigid state (fixed geometry), whereas in flexible models, they are allowed to undergo geometrical changes. Lastly there are complex models that include the effects of polarisation and many body effects – these models are useful for ionic systems and for modelling the interactions at the solute-solvent interface [3]. There are usually a large number of water molecules in a simulation, and as the complexity of the water model increases, the time needed for the simulation also increases, thus models including polarisation effects are usually used for more accurate work. Each model uses a suitable force field parameterised for the interactions between the water molecules. To test the accuracy of the model, simulations are run on water or solvated systems, from which thermodynamic or kinetic properties (such as enthalpies of vaporisation and diffusion coefficients) are calculated and then compared to experimental values [3, 26] – the model should give good agreement with experimental results.

In the simple models, the force field involves electrostatic and van der Waals interactions, which take the form of the Coulombic and Lennard-Jones potentials, respectively [3] (see Equations **2.16** – **2.18**). In the TIP3P [27] and SPC [28] models, the water molecules interact using three sites, located on the three atoms of each molecule. The electrostatic interactions involve both hydrogen and oxygen atoms (which have partial positive and partial negative charges, respectively), while the van der Waals interactions are calculated *only* between the oxygen atoms of a pair of molecules. The SPC model was later reparameterised to the improved SPC/E model [29], which gives a better density, radial distribution function and diffusion constant when compared with experiment [29]. These models are depicted in **Figure 2.3**. Four-site models, such as the TIP4P model [27], shift the negative charge on the oxygen atom towards the hydrogens to a point on the line bisecting the HÔH angle (marked M in **Figure 2.3**). The TIP4P model has been shown to give very good agreement with thermodynamic and structural experimental data for water [27]. In the TIP5P model [30], the negative charge is placed on the two lone pair sites of the oxygen atom (with each lone pair being assigned half

the negative charge), resulting in a five-site model – the lone pair sites are marked L in **Figure 2.3**. The TIP5P model shows general improvement over previous interaction-site models such as TIP3P and TIP4P – for example, the dielectric constant and temperature dependence of density calculated from the TIP5P model agree better with experimental values [30]. In the four- and five-site models, the electrostatic interactions are calculated using the charge sites in each model, and the van der Waals interactions are calculated in the same way as for the three-site models. In all models, the negative charges are balanced by the positive charges, so that the water molecule as a whole is neutral.

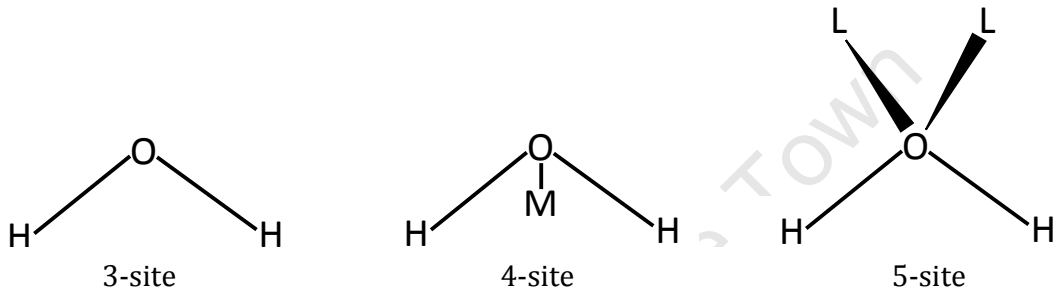


Figure 2.3: Schematic diagram of the 3-, 4- and 5-site water models, adapted from reference [3].

The geometrical and physical properties of the different water models are summarised in **Table 2.1**, where r, θ and q represent distance, angle and charge, respectively, and A and C are the Lennard-Jones parameters (Equation **2.17**) for the oxygen-oxygen interactions.

Table 2.1: Geometrical parameters and properties of different water models.

	SPC	SPC/E	TIP3P	TIP4P	TIP5P
r**(OH), Å**	1.0	1.0	0.9572	0.9572	0.9572
θ_{HOH}**, degrees**	109.47	109.47	104.52	104.52	104.52
A **x** 10^{-3}**, kcal Å**^{12}**/mol**	629.4	629.4	582.0	600.0	544.5
C**, kcal Å**^{6}**/mol**	625.5	625.5	595.0	610.0	590.3
q**(O)**	–0.82	–0.8476	–0.834	0.0	0.0
q**(H)**	0.41	0.4238	0.417	0.52	0.241
q**(M)**	–	–	–	–1.04	–
q**(L)**	–	–	–	–	–0.241
r**(OM), Å**	–	–	–	0.15	–
r**(OL), Å**	–	–	–	–	0.70
θ_{LOL}**, degrees**	–	–	–	–	109.47

* The values in the **Table 2.1** were taken from the following references: SPC [27, 29], SPC/E [29], TIP3P and TIP4P [27, 30], TIP5P [30].

2.4.5 Periodic Boundary Conditions

When simulating bulk liquids or large solvated systems, the simulation conditions (such as temperature, pressure, density) should be the same as if it were being done experimentally. Also, the particles in the simulation should experience the same forces and environment as they would in bulk solution. The solute and solvent particles are first placed in a simulation box or cell, where the box dimensions and number of solvent molecules are chosen appropriately so as to represent the correct solution density. However, if just one simulation box is used, molecules on the surface or edges of the box would experience different forces and interactions than those within the box (bulk solution) [1]. To solve this problem, *periodic boundary conditions* (PBC) are used, where the simulation box and its contents are repeated infinitely in three-dimensional (3D) space to form a space-filling periodic array. **Figure 2.4** is a two-dimensional (2D) illustration of PBC for a cubic cell, the simplest shape to use for the simulation box.

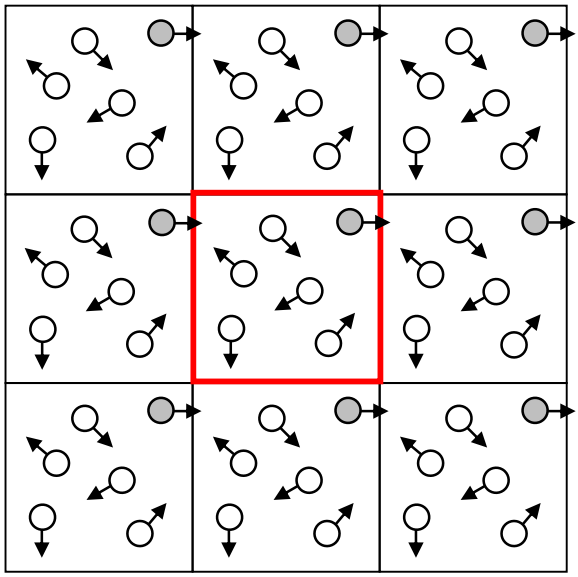


Figure 2.4: 2D representation of PBC, adapted from reference [3].

In two dimensions, the central cell is surrounded by 8 identical cells or images of itself (in 3D space, each cell is surrounded by 26 others). During the simulation, as a particle moves in the central box, all of its image particles move in the same way (illustrated by the shaded particles in **Figure 2.4**), and when a particle leaves the central box (highlighted in red) from one side, one of its image particles in an adjacent cell moves in from the opposite side with the same velocity to replace it [12]. Thus PBC ensure that the particles on the boundaries of the simulation box experience the same forces and interactions as those within the box, i.e. that all particles behave as bulk solution. It also ensures that the number of particles in the cells remains constant [3]. **Figure 2.5** depicts a cell (α-CD in water) from a simulation run for this thesis.

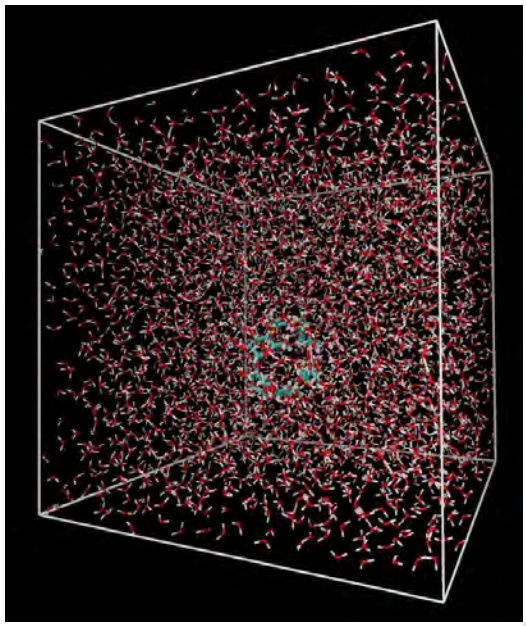


Figure 2.5: A cubic simulation box with an α-CD molecule surrounded by 4040 water molecules (from a simulation run in this thesis), obtained from the Visual Molecular Dynamics (VMD) application [31].

Besides the cubic cell, other shapes can also be chosen for the simulation box, provided the shape fills all space when it is replicated in three dimensions – examples include the rhombic dodecahedron [32], truncated octahedron [33, 34], and the hexagonal prism. Certain shapes may also be more suited to studying certain systems – for example, the truncated octahedron has a nearly spherical shape and may be better to use when simulating spherical molecules, while the hexagonal prism would be more suited to the cylindrically-shaped DNA molecule [3].

2.4.6 Potential Truncation and the Minimum Image Convention

In theory, PBC means each atom interacts with the other atoms in its own cell as well as *all* the image atoms in the periodic array, i.e. there are an infinite number of non-bonded interactions – however, this would be very time-consuming to evaluate, as the calculation of non-bonded energies is the most computationally-intensive part of a simulation [3]. Also, using PBC for liquids artificially treats the system like a crystal, and calculating all the non-bonded energies would reinforce the idea of long-range order and structure, which is not true for bulk liquids or solutions [12]. In addition, for systems with short-range non-bonded interactions (i.e. those that decrease rapidly with distance), it is only necessary to calculate the important interactions with *neighbouring*

atoms [3]. The above considerations give rise to the *minimum image convention* [12], where each atom interacts only with the nearest atoms or image atoms. This means:

1. The atom does not interact with any of its own images.
2. The atom does not interact simultaneously with another atom and one of the images of that atom [12].

The minimum image convention is illustrated for the cubic lattice in **Figure 2.6** [3], where the black atom interacts only with other atoms or images within the dashed line, which marks an imaginary cell centred on the black atom. This greatly reduces the number of non-bonded interactions that need to be calculated for each atom. Usually, a further approximation, known as a spherical cutoff, is also used. In **Figure 2.6**, the spherical cutoff is marked by the circle (sphere in three dimensions) with radius r_c (cutoff distance) drawn within the dashed line, where $r_c \leq ½L$ and L is the length of the simulation box. The non-bonded potential is truncated or set equal to zero for atoms lying outside the spherical cutoff (i.e. for $r \geq r_c$), thus the black atom only has non-bonded interactions with the atoms shaded in grey. Using a spherical cutoff saves time in the calculation of potential energies at each time step [1].

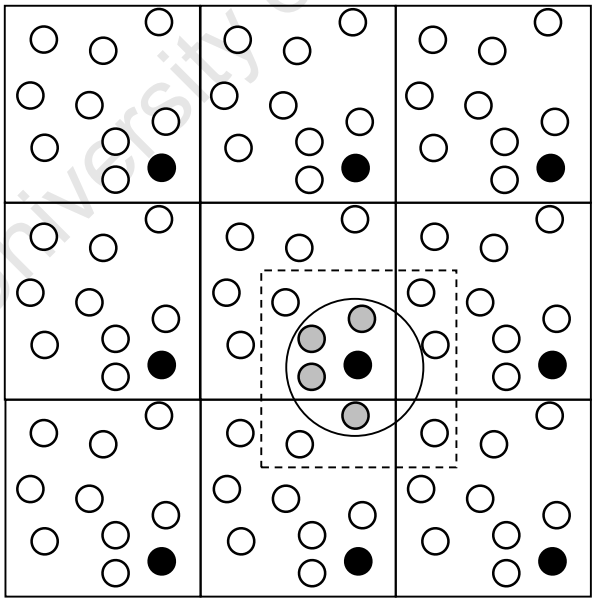


Figure 2.6: 2D representation of the minimum image convention and spherical cutoff, adapted from reference [3].

The size of the simulation box and the cutoff distance should be large enough so that the main non-bonded interactions are calculated, to ensure that the calculation of potential

energy is accurate. Thus for systems with long-range non-bonded interactions (for example, the long-range electrostatic interactions present in ionic systems), using a small spherical cutoff causes errors in the energy calculations. This can be solved by using a larger simulation box thereby increasing the cutoff distance, or by using other methods such as Ewald summation [1, 35].

In principle, to calculate the non-bonded interactions in a system, the distances between *all* pairs of non-bonded atoms would need to be calculated, to see which pairs are separated by less than the spherical cutoff [3]. However, this would be very time-consuming. To simplify the problem, a small neighbourhood region around the spherical cutoff is defined, with the coordinates of atoms within the cutoff and neighbourhood being stored in a *non-bonded neighbour* list [36]. In **Figure 2.7**, the spherical (r_c) and neighbour (r_n) cutoffs around the central black atom are indicated by solid and dotted lines, respectively, where $r_n > r_c$, and only the atoms shaded in grey and red are included in the neighbour list.

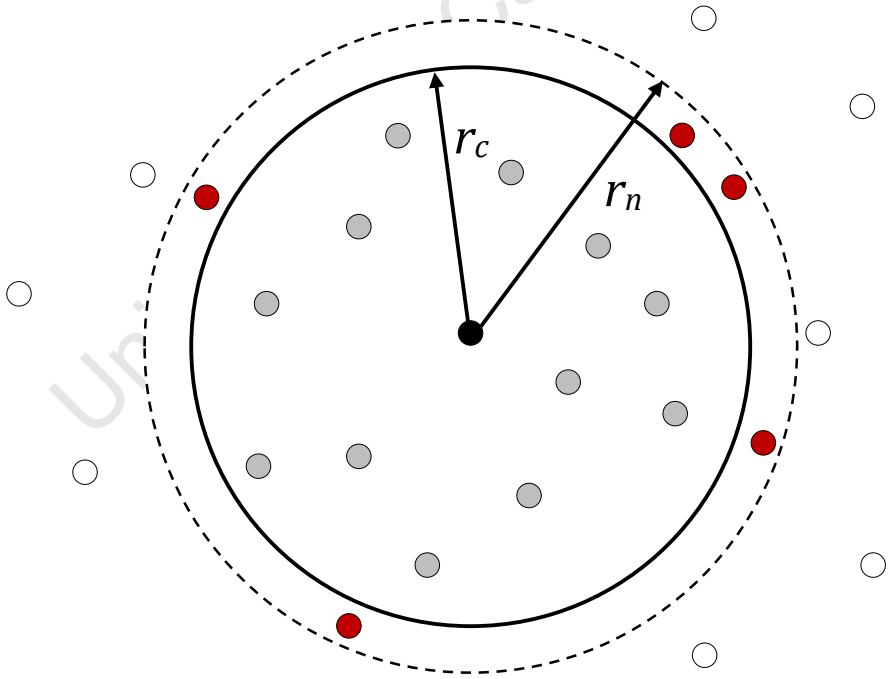


Figure 2.7: Schematic illustration of the spherical cutoff, neighbour cutoff and non-bonded neighbour list.

At each time step during the simulation, only the distances between an atom and those in its neighbour list are calculated, and the non-bonded interactions are then calculated

for distances within the spherical cutoff. The neighbour list is updated regularly during the simulation. The update frequency and size of the neighbourhood region should be chosen carefully, so that atoms outside the neighbour cutoff (white atoms in **Figure 2.7**) do not move within the spherical cutoff before the neighbour list is updated, as this would lead to errors in the energy calculations [3].

Another problem is that when the non-bonded interaction is abruptly set to zero at the cutoff radius, it causes a discontinuity in the interaction function – the total energy of the system has to be conserved, thus if the non-bonded potential energies suddenly become zero, the kinetic energy of the atoms (and therefore the system's temperature) will increase artificially to compensate [12]. To combat this problem, *shifting* or *switching* functions can be used to smoothen the transition of the non-bonded energy to zero at r_c.

In the first approach, the part of the interaction function before r_c is shifted towards the x-axis such that the interaction becomes zero at $r = r_c$, and for distances beyond r_c, the interaction is set to zero – this is called the *shifted* potential. Equation **2.24** below [3] is an example of a shifted potential $v'(r)$, where the original interaction potential $v(r)$ is shifted upwards by subtracting two terms from it, namely the constant potential v_c at r_c and a linear term involving the derivative of the interaction potential – including the linear term ensures that the derivative of the shifted potential at r_c is zero so that there are no discontinuities in the potential at r_c [3].

$$v'(r) = \begin{cases} v(r) - v_c - \left(\frac{dv(r)}{dr}\right)_{r=r_c} (r - r_c) & r \leq r_c \\ 0 & r > r_c \end{cases} \qquad \textbf{(2.24)}$$

Another approach is to define a small interval before the cutoff radius and multiply the non-bonded interaction function by a *switching* function over this interval. One such switching function is given in Equation **2.25** [12], where the lower limit of the interval (cuton distance) is denoted by r_l, and the upper limit is the cutoff distance r_c. The switching function $S(r)$ is equal to 1 before r_l and equal to 0 after r_c, thus it only changes

the interaction function in the interval (r_l, r_c), ensuring the potential changes smoothly to zero over this interval.

$$S(r) = \begin{cases} 1 & r \leq r_l \\ \dfrac{(r_c - r)^2(r_c + 2r - 3r_l)}{(r_c - r_l)^3} & r_l < r < r_c \\ 0 & r \geq r_c \end{cases} \qquad \textbf{(2.25)}$$

where $\dfrac{\mathrm{d}S}{\mathrm{d}r}(r_l) = \dfrac{\mathrm{d}S}{\mathrm{d}r}(r_c) = 0$

2.4.7 Simulation Ensembles

A molecular system can be characterised by its mechanical (microscopic) and thermodynamic (macroscopic) states. The mechanical state is defined by the instantaneous positions and momenta of all the atoms in the system, where each atom would be characterised by six coordinates (since both position and momentum are 3D vectors). For a particular configuration of N atoms, there would be $6N$ coordinates – these are often regarded as a single point in a multidimensional space known as phase space. On the other hand, the system's thermodynamic state is defined by a set of parameters such as the temperature (T), pressure (P), and number of atoms (N), which are held constant throughout a simulation. In a MD simulation, several different configurations of the system are generated – each of these configurations would have a different mechanical state (specified by different points in phase space), but all the configurations would have the same thermodynamic state. A collection of points in phase space with the same thermodynamic state is termed an *ensemble* [1].

MD simulations can be carried out in a variety of ensembles, some of the more commonly-used ones being: the microcanonical (constant-*NVE*) ensemble, where N, volume (V) and energy (E) are held fixed throughout the simulation; the canonical (constant-*NVT*) ensemble, where N, V, and T are fixed; the isothermal-isobaric (constant-*NPT*) ensemble, where N, P, and T are held constant; and the grand canonical (constant-μVT) ensemble, where the chemical potential (μ), V, and T are fixed [1]. The simulations in this thesis were run using the constant-*NPT* ensemble.

2.4.8 SHAKE Constraint Dynamics

In MD, it is important to choose the length of the time step appropriately to ensure both an accurate and efficient simulation (shorter time steps allow for more accurate simulations but also increase the computational time and resources). The time step must be chosen small enough to incorporate the highest frequency motions in the system, thus one can increase the length of the time step by constraining certain bond lengths or bond angles corresponding to high frequency motions / vibrations [12]. It is also important that the motion of the molecule as a whole not be altered too much when applying constraints. A very popular constraint method is the iterative SHAKE method [37], which is very useful for large systems such as macromolecules [12]. SHAKE is most often used to constrain the bonds between hydrogen (a light atom) and heavier atoms, as these vibrate with very high frequencies [3]. The main conformational changes in the system are usually low frequency motions, thus the less important high frequency motions (involving hydrogen for example) can be neglected [14] during the simulations.

References

1. Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Oxford University Press Inc.: New York, 1987.
2. Karplus, M.; Petsko, G. A. *Nature* **1990**, *347*, 631.
3. Leach, A. R. *Molecular Modelling: Principles and Applications*; Addison Wesley Longman Limited: England, 1996.
4. Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *Journal of Computational Chemistry* **1983**, *4*, 187.
5. Weiner, P. K.; Kollman, P. A. *Journal of Computational Chemistry* **1981**, *2*, 287.
6. Scott, W. R. P.; Hünenberger, P. H.; Tironi, I. G.; Mark, A. E.; Billeter, S. R.; Fennen, J.; Torda, A. E.; Huber, T.; Krüger, P.; van Gunsteren, W. F. *The Journal of Physical Chemistry A* **1999**, *103*, 3596.
7. Smith, W.; Forester, T. R. *Journal of Molecular Graphics* **1996**, *14*, 136.
8. Green, N. J. B. *Quantum Mechanics 1: Foundations*; Oxford University Press: New York, 1997.
9. Jensen, F. *Introduction to Computational Chemistry*; John Wiley & Sons: Chichester, 1999.
10. Szabo, A.; Ostlund, N. S. *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*; Dover Publications: Mineola, New York, 1996.
11. Best, R., *Combined NMR and Simulation Study of Carbohydrate Linkage Dynamics*, in *Faculty of Science. Department of Computational Chemistry.* 2000, M.Sc. University of Cape Town: Cape Town, South Africa.
12. van Gunsteren, W. F.; Berendsen, H. J. C. *Angewandte Chemie International Edition in English* **1990**, *29*, 992.
13. Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D. *Journal of Computational Chemistry* **2010**, *31*, 671.
14. Gamieldien, M. R., *Investigations of the Solution Properties of both Unmodified and Modified Cyclodextrins*, in *Faculty of Science. Department of Computational Chemistry.* 2008, M.Sc. University of Cape Town: Cape Town, South Africa.
15. Verlet, L. *Physical Review* **1967**, *159*, 98.

16. Hockney, R. W. *Methods in Computational Physics* **1970**, *9*, 136.
17. Kuttel, M.; Brady, J. W.; Naidoo, K. J. *Journal of Computational Chemistry* **2002**, *23*, 1236.
18. Ha, S. N.; Giammona, A.; Field, M.; Brady, J. W. *Carbohydrate Research* **1988**, *180*, 207.
19. Foley, B. L.; Tessier, M. B.; Woods, R. J. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2012**, *2*, 652.
20. Palma, R.; Himmel, M. E.; Liang, G.; Brady, J. W. In *ACS Symposium Series: Glycosyl Hydrolases in Biomass Conversion*; Himmel, M. E., Ed.; American Chemical Society: Washington, DC, 2001, p 112.
21. Murcko, M. A.; DiPaola, R. A. *Journal of the American Chemical Society* **1992**, *114*, 10010.
22. Guvench, O.; Greene, S. N.; Kamath, G.; Brady, J. W.; Venable, R. M.; Pastor, R. W.; Mackerell, A. D. *Journal of Computational Chemistry* **2008**, *29*, 2543.
23. Guvench, O.; Hatcher, E.; Venable, R. M.; Pastor, R. W.; MacKerell, A. D. *Journal of Chemical Theory and Computation* **2009**, *5*, 2353.
24. Guvench, O.; Mallajosyula, S. S.; Raman, E. P.; Hatcher, E.; Vanommeslaeghe, K.; Foster, T. J.; Jamison, F. W.; MacKerell, A. D. *Journal of Chemical Theory and Computation* **2011**, *7*, 3162.
25. Hatcher, E. R.; Guvench, O.; MacKerell, A. D. *Journal of Chemical Theory and Computation* **2009**, *5*, 1315.
26. Horn, H. W.; Swope, W. C.; Pitera, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.; Head-Gordon, T. *The Journal of Chemical Physics* **2004**, *120*, 9665.
27. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *The Journal of Chemical Physics* **1983**, *79*, 926.
28. Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J. In *Intermolecular forces*; Pullman, B., Ed. Reidel, Dordrecht, 1981, p 331.
29. Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. *The Journal of Physical Chemistry* **1987**, *91*, 6269.
30. Mahoney, M. W.; Jorgensen, W. L. *The Journal of Chemical Physics* **2000**, *112*, 8910.
31. Humphrey, W.; Dalke, A.; Schulten, K. *Journal of Molecular Graphics* **1996**, *14*, 33.
32. Wang, S.-s.; Krumhansl, J. A. *The Journal of Chemical Physics* **1972**, *56*, 4287.

33. Adams, D. J. *Chemical Physics Letters* **1979**, *62*, 329.
34. Adams, D. J. In *NRCC Proceedings: The Problem of Long-Range Forces in the Computer Simulation of Condensed Media*; Ceperely, D., Ed. Menlo Park, California, 1980; Vol. 9, p 13.
35. Ewald, P. *Ann. Phys.* **1921**, *64*, 253.
36. Verlet, L. *Physical Review* **1968**, *165*, 201.
37. Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. *Journal of Computational Physics* **1977**, *23*, 327.

Chapter 3

Computational Methods of Analysis

The conformational dynamics of carbohydrates in solution can be studied through statistical mechanics techniques such as time correlation functions. The local structuring of the solvent molecules around the carbohydrate moieties can also be studied through different techniques such as the calculation of pair and spatial distribution functions, as well as hydration numbers.

3.1 Background

The pair distribution function (PDF) – also known as the radial distribution function – is formally defined as the probability of finding a particular pair of atoms at a distance *r* apart, compared to the probability expected for a totally *random* distribution of the atoms (i.e. an ideal gas) at the same density [1]. When PDFs are applied to solvated systems, they give insight into the liquid structuring. In an MD study of hydrated CDs for example, PDFs were calculated from the trajectories to show the distribution of water molecules in the CD cavity as a function of their distance from the centres of mass of the CD rings [2]. In another study of CDs, PDFs were used to calculate the distribution of water molecules around the oxygen atoms of the hydroxyl groups on the glucose monomer units [3], giving information about the solvation shells around the CDs. PDFs however assume the same behaviour in all directions and give the average probability at each radial distance *r*. Thus to study the anisotropic (directionally dependent) structuring of liquids, it is more suitable to use non-radially averaged methods such as the spatial distribution function (SDF) [4]. The SDF is a three-dimensional (3D) plot of the probability density of water molecules around a solute molecule [4], and thus gives a detailed picture of the locations around the solute where it is more likely to find water molecules. For example, SDFs calculated from the MD trajectories of hydrated CDs revealed the local water probability density around the CDs, which was then used to compare the relative ordering (or structuring) imposed on the surrounding water by the CDs [5]. Other techniques to study the behaviour of water involve the calculation of residence times and hydration numbers. The residence time is defined as the time spent by water molecules in the first coordination shell of the solute before properly

reentering the bulk solution [6], and the hydration number is the number of incompressible water molecules in the first solvation shell of the solute [7]. Hydration numbers can be defined and obtained both experimentally [7-9] and computationally [6].

The conformational behaviour of the carbohydrate molecules themselves can also be studied. A useful technique from statistical mechanics is the time correlation function, which can give insight into the time-dependent conformational motions of molecules – this information is useful as it gives insight into the dynamical behaviour of molecules, which can be used to help explain their physical and chemical properties. Time correlation functions are used in this thesis and are discussed in more detail below.

3.2 Time Correlation Functions

Correlation functions are often used to measure how closely two quantities *x* and *y* are correlated with each other, and have been used extensively in the analysis of chemical systems. A simple correlation function is defined in Equation **3.1** below, where properties *x* and *y* of each atom *i* are multiplied together, then averaged over the total number of atoms M in the system to give the final correlation value C [10]. Positive or negative correlation values indicate some degree of correlation between the quantities, whereas a correlation of zero means the quantities are completely uncorrelated (independent) of each other. Normalising the function (dividing by the product of the root mean squares of *x* and *y*) ensures the correlation values lie between -1 and +1, where values close to these limits indicate high correlation [10, 11].

$$C_{xy} = \frac{1}{M}\sum_{i=1}^{M} x_i y_i \qquad (\mathbf{3.1})$$

In a dynamics simulation, it is often useful to calculate the correlation between two quantities at different *times* during the simulation, as this gives insight into *how* the two quantities lose correlation over time and the *rate of decay* of the correlation – this in turn can provide information on the *dynamical* behaviour of systems (rather than a *static* representation) which can help to explain processes taking place in real systems. The *Time Correlation Function* (TCF) is used to calculate the correlation between two quantities at different times during a simulation [10]. Equation **3.2** [10] shows a simple

TCF, where property x at an initial time (t=0) is correlated with property y at a later time t, where the brackets ⟨ ⟩ indicate averaging over all the time origins (i.e. initial times) used. TCFs are commonly used to analyse MD trajectories, since these trajectories represent changes in the system's configurations *over time*.

$$C_{xy}(t) = \langle x(t)y(0)\rangle \qquad (\mathbf{3.2})$$

If x and y are different quantities, the TCF is called a *cross-correlation function*, and if x and y represent the same quantity, the TCF is termed the *autocorrelation function* [10]. Various motions in the system (such as conformational changes, rotations, etc.) as well as transport processes can be analysed using TCFs. Transport properties such as the self-diffusion coefficient, viscosity and thermal conductivity can all be expressed as time integrals over certain TCFs [12]. For example, the self-diffusion coefficient can be written as an integral over the velocity autocorrelation function (Equation **3.3**) [12]. This type of analysis was started in the 1950's by Green [13, 14], who was the first to derive expressions for transport coefficients (such as diffusion and shear and bulk viscosity) in terms of TCFs.

$$D = \frac{1}{3}\int_0^\infty \langle \mathbf{v}(0)\cdot\mathbf{v}(t)\rangle\, dt \qquad (\mathbf{3.3})$$

Another useful application of autocorrelation functions in particular is that their Fourier transforms are related to experimental absorption spectra [1]. For instance, the infrared spectrum of a liquid is the Fourier transform of the dipole moment autocorrelation function [15]. Another example involves orientational autocorrelation functions, which measure how the rotations and reorientations of molecules change over time – these functions can often be related to different experimental spectra such as NMR, IR, and Raman spectra [10].

For the purposes of this thesis, autocorrelation functions will be further discussed. A schematic illustration of an autocorrelation function is given in **Figure 3.1**. For simplicity, suppose a MD simulation is run for 10 time steps of length Δt, generating 10 configurations (or frames), where $A(t_i)$ represents a property of the system measured for the configuration at every time step. In (i), the correlation is done for values of A one time step apart (i.e. Δt apart) – similarly in (ii) and (iii), the correlation is done for two and five time steps apart, respectively.

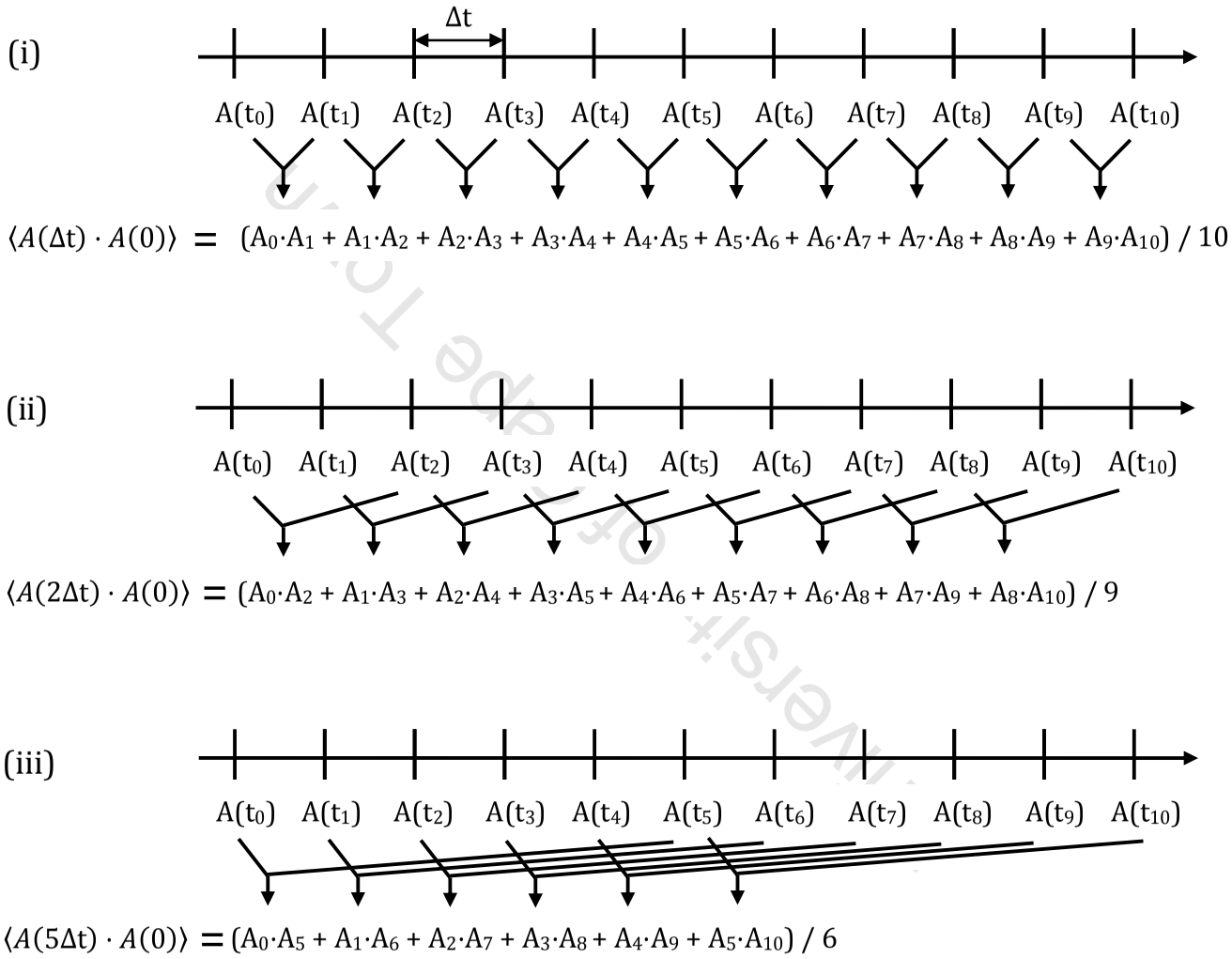


Figure 3.1: Schematic representation of an autocorrelation function (not normalised), adapted from reference [16].

TCFs can also be normalised by dividing by the average of the square of the time origins used – the normalised autocorrelation function is given below (adapted from reference [10]):

$$C_{AA}(t) = \frac{\langle A(t)A(0)\rangle}{\langle A(0)A(0)\rangle} \tag{3.4}$$

In the normalised function, the initial correlation between the property A, at each time step, and itself is uniform resulting in a correlation of 1, that is:

$$\text{At } t = 0, \qquad C_{AA}(0) = \frac{\langle A(0)A(0)\rangle}{\langle A(0)A(0)\rangle} = 1$$

Then as t increases, the correlation between frames decreases and eventually approaches zero as $t \to \infty$, i.e. the frames become uncorrelated [10]. The correlation often decreases exponentially with time, in which case a plot of $C_{AA}(t)$ versus t can be fit to the exponential decay function given in Equation **3.5**:

$$f(t) = \mathrm{A} + \mathrm{B}e^{-(t/\tau)} \tag{3.5}$$

where A and B are constants.

The rate at which the correlation decreases is measured by the *relaxation time* (or *correlation time*), which is represented by the quantity τ in Equation **3.6** – the smaller the value of τ, the faster the rate of decay. **Figure 3.2** is a graphical representation of a simple TCF following exponential decay. The TCF should be calculated as accurately as possible during the relaxation period, so that an accurate rate of decay is measured. One way to improve the accuracy is to increase the number of time origins used in the TCF calculation, which can be done by increasing the length of the simulation. In general, the simulation time should be much longer than the relaxation time [10]. The beginning correlation values are the most accurate as they are calculated using more data points than correlation values obtained for frames further apart. Typically, TCF values are usually only reliable up to about 1/4 to 1/3 of the trajectory length [17]. For this reason, typically only the beginning portions of the TCF plot are fitted to the exponential function when calculating the relaxation times.

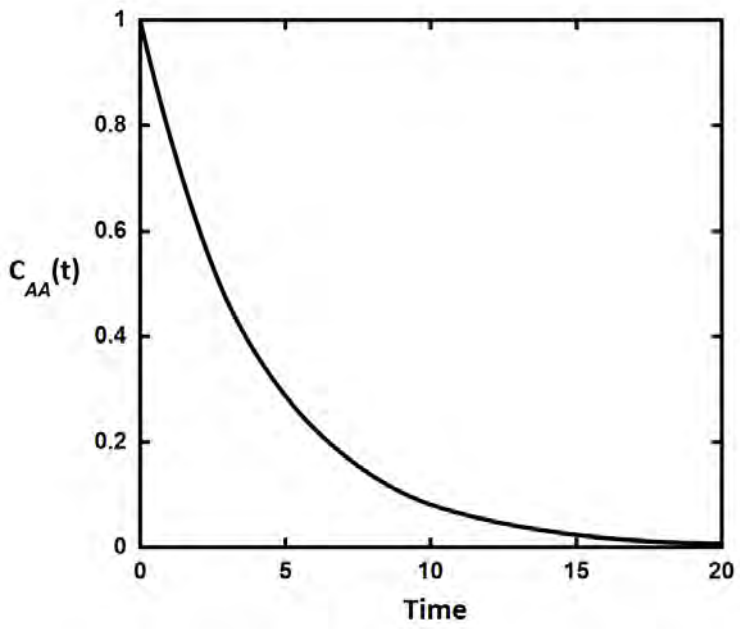


Figure 3.2: Schematic graph of a TCF following exponential decay.

An alternate way to define the autocorrelation function involves correlating the *deviations* (or *fluctuations*) in $A(t)$ from the equilibrium average value, rather than the $A(t)$ values themselves [18]. The equilibrium average $\langle A \rangle$ is the average value of property A measured over the simulation (Equation **3.6**), and is independent of time. N in Equation **3.6** is the total number of frames in the trajectory. The deviation $\delta A(t)$ and the corresponding normalised correlation function $C_{AA}(t)$ are defined in Equations **3.7** and **3.8** [18]. As before, $C_{AA}(0) = 1$ and $C_{AA}(t) \to 0$ as $t \to \infty$. This form of the TCF is used for the work in this thesis.

$$\langle A \rangle = \frac{1}{N} \sum_{i=1}^{N} A_i \qquad \textbf{(3.6)}$$

$$\delta A(t) = A(t) - \langle A \rangle \qquad \textbf{(3.7)}$$

$$C_{AA}(t) = \frac{\langle \delta A(t) \delta A(0) \rangle}{\langle \delta A(0) \delta A(0) \rangle} \qquad \textbf{(3.8)}$$

Figure 3.2 depicts an "idealised" TCF, where the function decays smoothly to zero as the time approaches infinity. In real simulations however, the TCF decays towards zero during the relaxation time, but as the time progresses, small-amplitude peaks often

show up regularly about the zero-correlation line (i.e. the correlation fluctuates between positive and negative values), where the peaks represent some type of periodic or regular motion/s that are present in the system. These regular motions are responsible for the decay of the TCF [19] and can be analysed further by calculating the Fourier transform of the TCF. Equation **3.9** below gives the Fourier transform $\hat{C}(\omega)$ of an autocorrelation function $C(t)$ [1].

$$\hat{C}(\omega) = 2\int_0^\infty \mathrm{d}t\; C(t)\cos\omega t \qquad (\mathbf{3.9})$$

The Fourier spectrum is a plot of intensity versus frequency, where the frequency peaks in the spectrum correspond to the frequencies of the regular motions observed in the TCF [19]. Thus the Fourier spectrum effectively breaks up the overlapped motions in the TCF into separate frequency signals – the number of frequency peaks gives an idea of how many different motions are present, while the relative intensity of the peaks indicates which are the more prominent / major motions.

References

1. Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Oxford University Press Inc.: New York, 1987.
2. Raffaini, G.; Ganazzoli, F. *Chemical Physics* **2007**, *333*, 128.
3. Perez-Miron, J.; Jaime, C.; Ivanov, P. M. *Chirality* **2008**, *20*, 1127.
4. Gamieldien, M. R., *Investigations of the Solution Properties of both Unmodified and Modified Cyclodextrins*, in *Faculty of Science. Department of Computational Chemistry.* 2008, M.Sc. University of Cape Town: Cape Town, South Africa.
5. Naidoo, K. J.; Chen, J. Y.-J.; Jansson, J. L. M.; Widmalm, G.; Maliniak, A. *The Journal of Physical Chemistry B* **2004**, *108*, 4236.
6. Impey, R. W.; Madden, P. A.; McDonald, I. R. *The Journal of Physical Chemistry* **1983**, *87*, 5071.
7. Junquera, E.; Olmos, D.; Aicart, E. *Physical Chemistry Chemical Physics* **2002**, *4*, 352.
8. Shikata, T.; Takahashi, R.; Satokawa, Y. *The Journal of Physical Chemistry B* **2007**, *111*, 12239.
9. Burakowski, A.; Gliński, J. *Chemical Physics* **2007**, *332*, 336.
10. Leach, A. R. *Molecular Modelling: Principles and Applications*; Addison Wesley Longman Limited: England, 1996.
11. Jensen, F. *Introduction to Computational Chemistry*; John Wiley & Sons: Chichester, 1999.
12. McQuarrie, D. A. *Statistical Mechanics*; University Science Books: Sausalito, California, 2000.
13. Green, M. S. *The Journal of Chemical Physics* **1952**, *20*, 1281.
14. Green, M. S. *The Journal of Chemical Physics* **1954**, *22*, 398.
15. Guillot, B. *The Journal of Chemical Physics* **1991**, *95*, 1543.
16. Kofke, D. A., *"Evaluating Time Correlation Functions"*, available from http://www.eng.buffalo.edu/~kofke/ce530/Lectures/Lecture12/sld012.htm, accessed on 17 September 2012.
17. *"CHARMM c33b2 correl.doc"*, available from http://www.charmm.org/documentation/c33b2/correl.html, accessed on 17 September 2012.

18. Chandler, D. *Introduction to Modern Statistical Mechanics*; Oxford University Press Inc.: New York, 1987.
19. Best, R., *Combined NMR and Simulation Study of Carbohydrate Linkage Dynamics*, in *Faculty of Science. Department of Computational Chemistry.* 2000, M.Sc. University of Cape Town: Cape Town, South Africa.

Chapter 4

Methods of Analysing Ring Conformations

The conformational behaviour of cyclic molecules plays an important role in their physical properties [1-4] and chemical reactivity [5-9] – this has been discussed in Chapter 1 of this thesis. The next step is to develop a *systematic* way to characterise the out-of-plane puckering motion of rings as well as to identify the conformers. There are a vast number of conformations for a ring system, which can be broken down into a small number of characteristic forms (often termed the *canonical* conformations [10]), and a vast number of intermediate forms. Systematic techniques have been developed to characterize ring puckers both *qualitatively* and *quantitatively*. The *qualitative* nomenclature of puckers is *discrete* in the sense that it is applied to the naming of the finite number of canonical forms and their states, while their *quantitative* characterisation involves developing *continuous* models that can be applied to both canonical and intermediate forms. In quantitative models, a reduced number of parameters are defined based on perpendicular displacements from a mean plane of the ring [11-15], the intracyclic torsion angles [16-22], or the triangular tessellation of the ring [10, 14, 23, 24]. These pucker classification systems are described in this chapter, with particular emphasis on quantitative models, one of which is the focus of this thesis.

4.1 *Qualitative* Characterisation of Puckers

Qualitatively, terms such as “chair” and “boat” have been used to describe the shapes of *canonical* conformations of rings [25]. A set of rules for the systematic naming of carbohydrates and their conformations was developed formally by the IUPAC-IUB Joint Commission on Biochemical Nomenclature in 1971, which has since been revised [26-29]. This system (here called the *IUPAC system*) has been widely used to characterise the conformations of 5-membered [29, 30] and 6-membered rings [6, 20, 31, 32], and has also been extended to the naming of larger 7-membered [29, 33-35] and 8-membered [34, 36] rings.

4.1.1 IUPAC Nomenclature of 6-membered Rings

There are five canonical conformations a 6-membered ring can take on, namely, the *chair* (*C*), *boat* (*B*), *half-chair* (*H*), *skew-boat* (*S*), and *envelope* (*E*) conformations [27]. We term these five canonical conformations *canonical shapes*, or simply *shapes*, as depicted in **Figure 4.1**. According to the IUPAC system [27], for each shape a reference plane is defined by certain ring atoms – this is called the *IUPAC plane*, and is shaded in light orange in **Figure 4.1.**

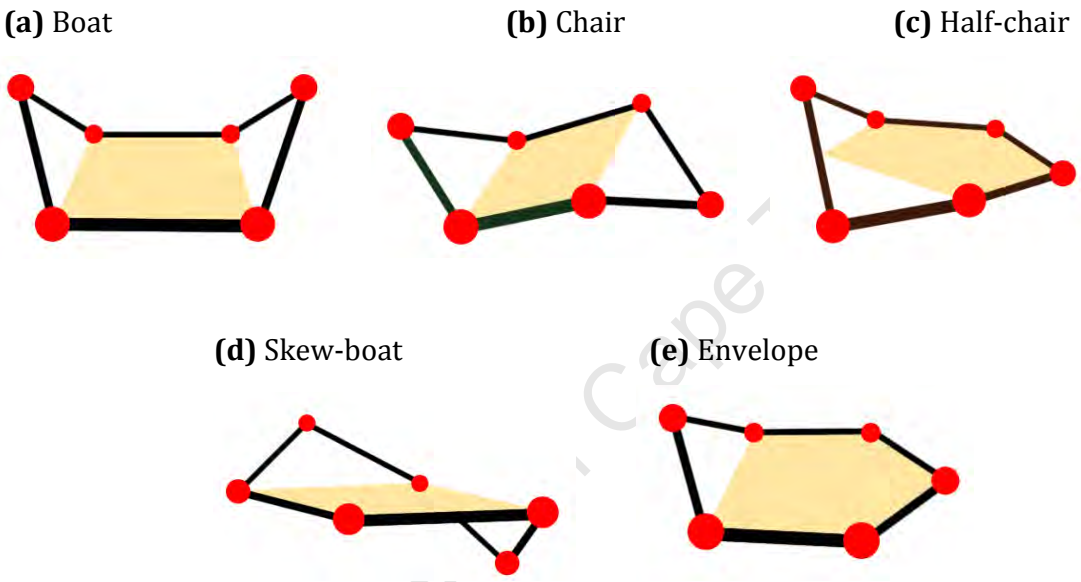


Figure 4.1: The five canonical conformations (or shapes) of 6-membered rings, adapted from reference [10]. The IUPAC plane has been shaded in light orange.

When classifying the puckers of 6-membered rings (and also *N*-membered rings in general, where $N \geq 3$), ***only*** the ring atoms are considered, that is the substituents on the ring atoms are not explicitly taken into account. In this section, we shall limit ourselves to explicitly discussing 6-membered rings. There are two types of rings to consider, namely ***homogeneous*** and ***heterogeneous*** rings. In homogeneous rings, all of the ring atoms are identical – an example is cyclohexane, where all of the ring atoms are carbon. In heterogeneous rings, the ring atoms are not all identical, an example being glucose (a pyranose), where the ring atoms include five carbon atoms and one oxygen atom. In classifying the puckers of 6-membered *homogeneous* rings, it may be sufficient to use the five canonical shapes to describe their shapes. However, this may not be sufficient if there are a number of ring substituents that reduce symmetry and change the

environment for a particular atom, thus altering the preferred ring conformation. On the other hand, for *heterogeneous* rings, each canonical shape gives rise to *different* conformations, which have that same basic shape (e.g. *boat* shape) but differ in the *positions of ring atoms relative to the IUPAC plane*. These *different* conformations of each shape are called ***states*** (for homogeneous systems, the states of each shape are equivalent). The IUPAC system is used to classify these states, and is described briefly below for 6-membered rings [27]:

1. The ring atoms are numbered from 1 to 6 in a clockwise manner (which will be referred to as ***IUPAC numbering*** here). The atom numbered 1 is arbitrary for homogeneous molecules, unless there are ring substituents which would impart a preferential ordering. For heterogeneous molecules, the ring ordering must be specified, where for pyranoses, the anomeric carbon is usually numbered 1.
2. For each ring shape, certain ring atoms are chosen to define the IUPAC plane according to a set of rules, i.e. the IUPAC plane differs depending on the shape [26].
3. An italicised capital letter denotes the ring shape, e.g. *C* denotes the *chair* shape.
4. Ring atoms that lie outside the IUPAC plane are either *above the plane* (the side of the plane from which the IUPAC numbering appears ***clockwise***) or *below the plane* (the side from which the numbering appears ***anti-clockwise***). Ring atoms *above* and *below* the plane are indicated by writing their IUPAC numbers as *superscripts* and *subscripts*, respectively.

As an example, the ${}^{4}C_{1}$ state of 6-membered rings has a *chair* shape, where atoms 2, 3, 5, and 6 define the IUPAC plane, with atom 4 located above this plane and atom 1 below the plane. The five shapes of 6-membered rings and their respective states (38 in total) are given in **Table 4.1** [10]. These states have been used to classify several types of heterogeneous 6-membered rings – examples include saturated heterocyclic rings with one or more ring heteroatoms such as oxygen, nitrogen or sulphur [23, 31, 37].

Table 4.1: The 38 states for the five canonical shapes of 6-membered rings [10].

Shape	Number of states	States
Chair	2	${}^{4}C_1$ ${}^{1}C_4$
Boat	6	${}^{1,4}B$ $B_{1,4}$ ${}^{2,5}B$ $B_{2,5}$ ${}^{3,6}B$ $B_{3,6}$
Half-chair	12	${}^{1}H_2$ ${}^{2}H_1$ ${}^{2}H_3$ ${}^{3}H_2$ ${}^{3}H_4$ ${}^{4}H_3$ ${}^{4}H_5$ ${}^{5}H_4$ ${}^{5}H_6$ ${}^{6}H_5$ ${}^{6}H_1$ ${}^{1}H_6$
Skew-boat	6	${}^{1}S_3$ ${}^{3}S_1$ ${}^{5}S_1$ ${}^{1}S_5$ ${}^{6}S_2$ ${}^{2}S_6$
Envelope	12	${}^{1}E$ E_1 ${}^{2}E$ E_2 ${}^{3}E$ E_3 ${}^{4}E$ E_4 ${}^{5}E$ E_5 ${}^{6}E$ E_6
	38 (total)	

4.2 *Quantitative* Characterisation of Puckers

Often cyclic molecules take on distorted conformations that are intermediate between the basic canonical forms. In these cases, qualitative descriptions are insufficient and quantitative ways to describe puckers become necessary. For example, puckers can be specified quantitatively using Cartesian coordinates (the x, y, z positional coordinates of all the atoms in the system in three-dimensional space), internal coordinates (coordinates specifying the positions of atoms *relative* to others by using for example, bond distances, bond angles, torsion angles), or using a system of puckering coordinates, which use a considerably reduced number of parameters, as compared to the larger number of internal or Cartesian coordinates (there are $3N$ Cartesian coordinates for a system of N atoms). To derive a system of puckering coordinates, the Cartesian or internal coordinates of the ring atoms *only* are related to a smaller number of parameters through mathematical relationships to generate a puckering model. There are always N-3 puckering coordinates, for an N-membered ring. Section 4.2.1 describes some advantages and general applications of using ring pucker coordinates.

4.2.1 Advantages of Ring Pucker Coordinates

Pucker coordinates are few in number (at least for 5- and 6-membered rings) and are simple to represent graphically, for example through two-dimensional (2D) plots or three-dimensional (3D) surfaces such as spheres or tori [12, 14, 35, 36]. These graphical representations are often called **conformational maps**, which show all possible canonical conformations of the ring and also help to visualise the conformations in relation to their puckering coordinates. **Figure 4.2** is an example of a conformational map for a pyranose system [29].

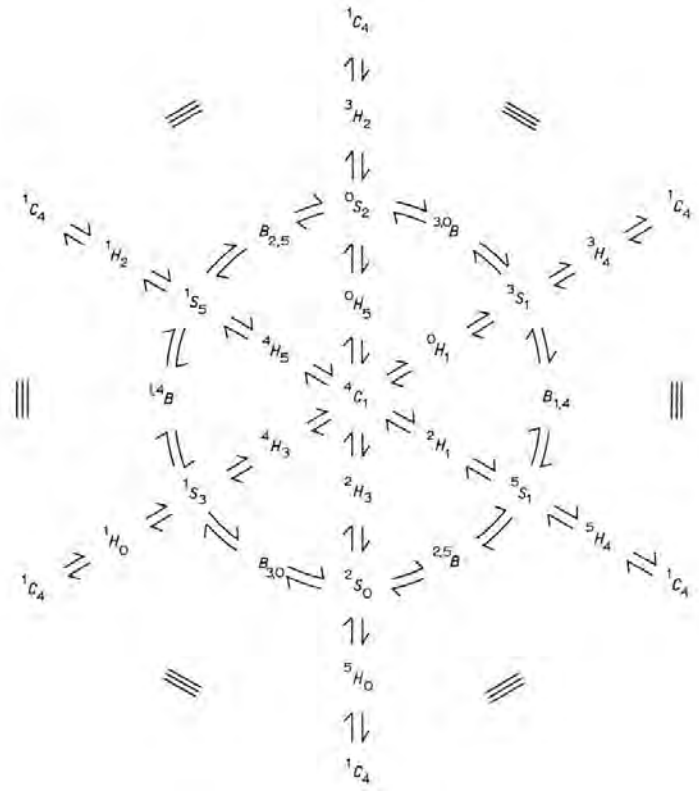


Figure 4.2: Conformational map showing the interconversion paths between conformations of a pyranose ring, from reference [29]. The letter "O" included as a superscript / subscript refers to the ring oxygen atom, which is atom 6 according to the IUPAC numbering.

From these conformational maps, inter-conversion paths between different conformers and changes in pucker coordinates along these paths can be identified – this has been done for several 6- to 8-membered rings [29, 35, 36, 38]. It is important to note that these inter-conversion paths only show *possible* conformational routes in moving from one conformer to another, they do not give any information regarding the relative stabilities of the conformers [6], and are *not necessarily* the conformational pathways observed in real systems (e.g. biochemical reactions).

Another application of puckering coordinates is to use the puckering models with potential energy or free energy functions to generate **conformational *energy or free energy* maps**, which graphically illustrate the relative energies (and therefore the relative stabilities) of the conformers either directly [5] or indirectly [6] as a function of their puckering coordinates. From these energy maps, it *is* possible to identify transition paths between conformers that are observed in real systems – this has several

advantages, such as gaining insight into the mechanisms of reactions and other chemical processes [5, 6].

Sections 4.2.2 to 4.2.4 provide an outline of some quantitative approaches used to characterise ring pucker, with the focus being on 5- to 8-membered rings. Examples of graphical representations of the pucker coordinates, as well as some applications to real systems will also be discussed. In all cases, *N* refers to the number of ring atoms.

4.2.2 Displacements from a Mean Plane

This approach involves defining a mean plane for the ring, then expressing the perpendicular out-of-plane displacements of the ring atoms from this plane in terms of puckering coordinates. The displacements are depicted schematically in **Figure 4.3**.

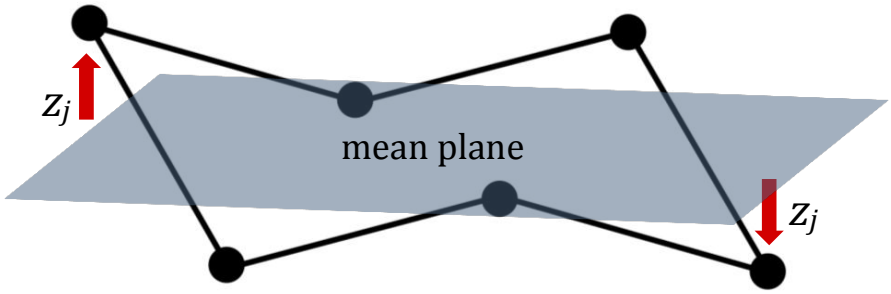


Figure 4.3: Schematic representation of a 6-membered ring, showing the perpendicular out-of-plane displacements z_j of the ring atoms from the mean plane (in light blue), labelled here for two of the atoms.

The first theory on ring puckering was proposed by Kilpatrick *et al.* in 1947 [11] to model the puckering of cyclopentane (a 5-membered ring). They expressed the perpendicular displacements z_j of the carbon atoms from the mean plane of the ring in terms of two puckering parameters q and ψ, as shown in Equation **4.1**:

$$z_j = \sqrt{2/5}\, q \cos\left(2\psi + \frac{4\pi(j-1)}{5}\right) \qquad \textbf{(4.1)}$$

where $1 \leq j \leq 5$ ($j \in \mathbb{Z}$), q is called the puckering amplitude, ψ is the phase angle. The puckers of cyclopentane have been effectively classified by two parameters where the puckering amplitude measures the extent to which the ring is puckered from the planar form, while the phase angle describes the shape of the ring [25].

A modified form of Equation **4.1** was also used to study the conformational behaviour of tetrahydrofuran (a heterocyclic 5-membered ring) [39]. There are however some limitations to Equation **4.1** – it can only be applied to *equilateral* 5-membered rings (i.e. with equal bond lengths and angles), and it does not correctly describe the geometry of a 5-membered ring since as the molecule undergoes conformational changes, the bond lengths change in a way that is not physically reasonable [40]. An improvement on this model was made by Adams *et al.* in 1970, who introduced certain constraints to Equation **4.1** to ensure that all carbon-carbon bond lengths remain constant as the cyclopentane molecule changed conformation [41]. A further improvement was made by Herzyk and Rabczenko [42], whose model both ensures that the carbon-carbon bond lengths remain constant and that the ring's centre of mass does not change position during the conformational changes.

Puckering models for *general N*-membered rings have also been developed. In 1971, Pickett and Strauss [13] used a group theoretical approach to describe the conformations of *N*-membered *cycloalkane* and *substituted cycloalkane* rings. Starting from the irreducible representations of the normal mode displacements of the D_{Nh} group (symmetry group for a planar cycloalkane molecule), they represented the out-of-plane displacements of a general ring conformation from the planar form as the linear combination Γ of these normal modes, as shown in Equations **4.2** and **4.3**.

When *N* is even: $$\Gamma = B_{2(g,u)} + E_{2u} + \ldots + E_{(N/2-1)(u,g)} \quad (\mathbf{4.2})$$

When *N* is odd: $$\Gamma = E_2'' + \ldots + E_{(N-1)/2}'' \quad (\mathbf{4.3})$$

For each representation in the expressions for Γ, z_j is written in terms of symmetry-adapted pucker) coordinates, so that the overall expression of z_j for a general conformation is given by Equations **4.4** and **4.5**. Here, z_j measures the displacements from the *planar* ring (i.e. a *mean* plane is not explicitly defined), and $1 \leq j \leq N$ ($j \in \mathbb{Z}$).

When *N* is even:

$$z_j = (-1)^j Q + \sum_m \rho_m \cos\left(\frac{2\pi jm}{N} + \varphi_m\right) \quad (\mathbf{4.4})$$

where $m = 2, \ldots, N/2 - 1$.

When N is odd:

$$z_j = \sum_m \rho_m \cos\left(\frac{2\pi jm}{N} + \varphi_m\right) \qquad \textbf{(4.5)}$$

where m = 2, … , (N–1)/2. Q, φ_m and ρ_m are the puckering coordinates, herein referred to as Pickett-Strauss (PS) coordinates. 6-membered rings are characterised by three coordinates Q, φ_2 and ρ_2, and in general there are N-3 puckering parameters for an N-membered ring. Pickett and Strauss used these coordinates to obtain conformational energy expressions and proposed a method to locate possible inter-conversion paths between conformers.

In 1975, Bocian *et al.* [14] used the PS coordinates to describe the conformations of cycloheptane (a 7-membered ring) in terms of four pucker parameters. Setting N = 7 and m = 2, 3 in Equation **4.5**, the expression for z_j becomes:

$$z_j = \rho_2 \cos\left(\frac{2\pi j2}{7} + \varphi_2\right) + \rho_3 \cos\left(\frac{2\pi j3}{7} + \varphi_3\right) \qquad \textbf{(4.6)}$$

where ρ_2, ρ_3, φ_2, φ_3 are the pucker parameters. Then setting $\rho_2 = \rho \cos\theta$ and $\rho_3 = \rho \sin\theta$, Equation **4.6** was rewritten in terms of the parameters ρ, θ, φ_2, φ_3 as in Equation **4.7**:

$$z_j = \rho\left[\cos\theta \cos\left(\frac{2\pi j2}{7} + \phi_2\right) + \sin\theta \cos\left(\frac{2\pi j3}{7} + \phi_3\right)\right] \qquad \textbf{(4.7)}$$

where ρ reflects the extent of ring puckering, and the angles θ, φ_2, φ_3 are related to the pucker shape. Six canonical forms for cycloheptane were identified and characterised in terms of these parameters. Bocian *et al.* also introduced a way to visualise these coordinates, by representing them on a set of tori (see **Figure 4.4**), where points on the surface of the tori correspond to particular conformations [14].

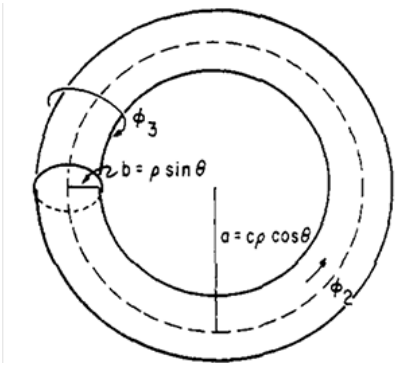


Figure 4.4: Graphical representation of the pucker coordinates ρ, θ, φ_2, φ_3 for cycloheptane, represented on a torus (taken from reference [14]).

Potential energies of cycloheptane conformers were calculated and used to characterise possible interconversion pathways and transition states. The pathways were represented pictorially on helices wound around the tori – for example, **Figure 4.5** shows the interconversion path between the *chair* and *twist-chair* conformers [14].

Figure 4.5: Interconversion path between the *chair* and *twist-chair* conformers of cycloheptane, taken from reference [14].

Evans and Boeyens [15] followed a group theoretical approach [13] to express the conformations of 5- to 8-membered rings as a linear combination of a few basic conformations or *primitive forms*. For each *N*-membered ring, they identified *N*-3 primitive forms, e.g. the three primitive forms for a 6-membered ring are the B_{2g}, E_{2u} (cos form), and E_{2u} (sin form) forms, which correspond to the *chair*, *boat*, and *twist-boat* conformations, respectively. A general conformation is expressed as a linear combination of the primitive forms, where the linear coefficients give the relative contributions of each primitive form. In this way, each conformation can be visualised as a combination of a few familiar basic shapes. Examples of 5- and 6-membered rings from literature, such as the furanoid and pyranoid rings of sucrose, have been analysed using this method [12, 15].

An alternative to the group theoretical approach is the *N*-membered ring pucker model proposed by Cremer and Pople [12, 43]. A mean plane through the *N* ring atoms is first defined, then the displacements z_j from this plane are used to define the puckering coordinates q_m (amplitudes) and Φ_m (phase angles), as shown in Equations **4.8** to **4.10**. These will be termed Cremer-Pople (CP) coordinates here, and in general, there are *N*-3 CP coordinates for an *N*-membered ring. The CP coordinates for 5- to 8-membered rings are given in **Table 4.2**.

$$q_m \cos \Phi_m = (2/N)^{\frac{1}{2}} \sum_{j=1}^{N} z_j \cos[2\pi m(j-1)/N] \qquad \textbf{(4.8)}$$

$$q_m \sin \Phi_m = -(2/N)^{\frac{1}{2}} \sum_{j=1}^{N} z_j \sin[2\pi m(j-1)/N] \qquad \textbf{(4.9)}$$

where $1 \leq j \leq N$ ($j \in \mathbb{Z}$), m = 2, 3, … , (N-1)/2 (when N is odd) and m = 2, 3, ... , ½N-1 (when N is even). When N is even, there is an additional puckering coordinate $q_{N/2}$ defined as:

$$q_{N/2} = N^{-\frac{1}{2}} \sum_{j=1}^{N} (-1)^{j-1} z_j \qquad \textbf{(4.10)}$$

Table 4.2: Cremer-Pople pucker coordinates for 5- to 8-membered rings.

Ring system	CP coordinates
5-membered ring	(q_2, Φ_2)
6-membered ring	(q_2, q_3, Φ_2)
7-membered ring	(q_2, q_3, Φ_2, Φ_3)
8-membered ring	(q_2, q_3, q_4, Φ_2, Φ_3)

The z_j can also be written in terms of the CP coordinates, allowing Cartesian coordinates to be calculated if the puckering coordinates are known. In particular when N = 5, the expression for z_j becomes:

$$z_j = \sqrt{2/5}\, q \cos\left(\Phi + \frac{4\pi(j-1)}{5}\right) \qquad \textbf{(4.11)}$$

where the subscripts '2' from q_2 and Φ_2 have been dropped. Equation **4.11** is the same expression obtained by Kilpatrick *et al.* [11] for cyclopentane, where q and Φ correspond to the parameters q and 2ψ in Equation **4.1**, respectively. However, the quantity Φ in Equation **4.11** also allows the CP model to be applied to a *general* 5-membered ring with different bond lengths and angles [12].

Cremer and Pople also gave a very useful graphical representation of the pucker coordinates for a 6-membered ring. For N = 6, the CP coordinates (q_2, q_3, Φ_2) can be replaced by spherical polar coordinates (Q, θ, Φ), where Q is the total puckering amplitude, and Φ and θ are angles in the intervals [0°, 360°] and [0°, 180°], respectively. Using these spherical coordinates for a constant value of Q, all the puckers of a 6-membered ring can be mapped onto the surface of a sphere (see **Figure 4.6**), which helps to visualise the conformations in relation to their pucker coordinates. In general,

it should also be possible to represent the puckers of an *N*-membered ring on the surface of a hypersphere in *N*-3 dimensions [12].

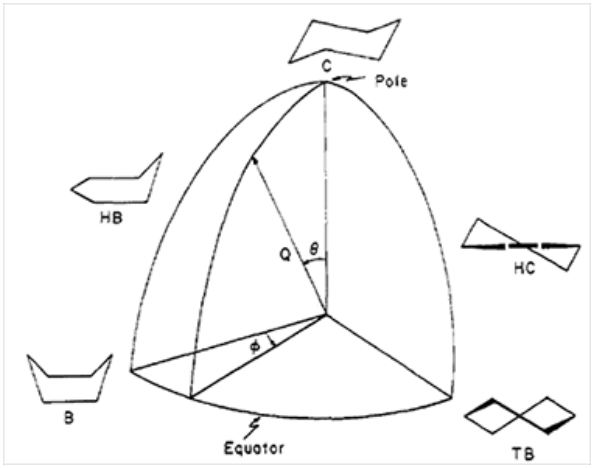


Figure 4.6: One octant of the sphere onto which puckers of 6-membered rings can be mapped, using the CP spherical coordinates (Q, θ, Φ) for constant Q (from reference [12]).

The CP coordinate system is exact [19], and can be applied to any cyclic molecule with no approximations, implying that the ring can have heteroatoms and arbitrary bond lengths and angles [12]. Conformations of several cyclic molecules in literature have been characterised with CP coordinates, some of which are outlined as follows.

CP coordinates of small-ring systems (N=5, 6)

CP coordinates are useful for molecular dynamics simulations. In particular, they have also been used as collective variables in metadynamics simulations [44, 45]. Sega *et al.* [45] use the CP spherical coordinates to characterise the puckers of the 6-membered ring, glucuronic acid, and to generate its pucker free energy surface using CP coordinates as collective variables. Bérces *et al.* also report CP spherical coordinates calculated for several substituted 6-membered rings [20].

CP coordinates of medium-ring systems (N>6)

Relatively fewer studies have been undertaken for medium sized *N*-membered rings where *N*>6. Boessenkool and Boeyens [35] used CP coordinates to characterise conformations of 7-membered rings, representing the coordinates on tori and generating maps of possible interconversion pathways between the conformers. Evans and Boeyens [36] presented a way to map the five CP coordinates of 8-membered rings

onto a 3D surface, which involves a set of tori placed on a sphere. The 3D surface is generated as follows: a unit sphere is defined, where the polar angle θ (defined in terms of the CP coordinates q_2, q_3, and q_4) measures the angular distance from the north pole of the sphere ($0 \leq \theta \leq \pi$). For each θ value, a torus is defined in terms of the CP coordinates q_2, q_3, Φ_2, Φ_3, and the torus is placed on the sphere at the corresponding value of θ – this is illustrated in **Figure 4.7**.

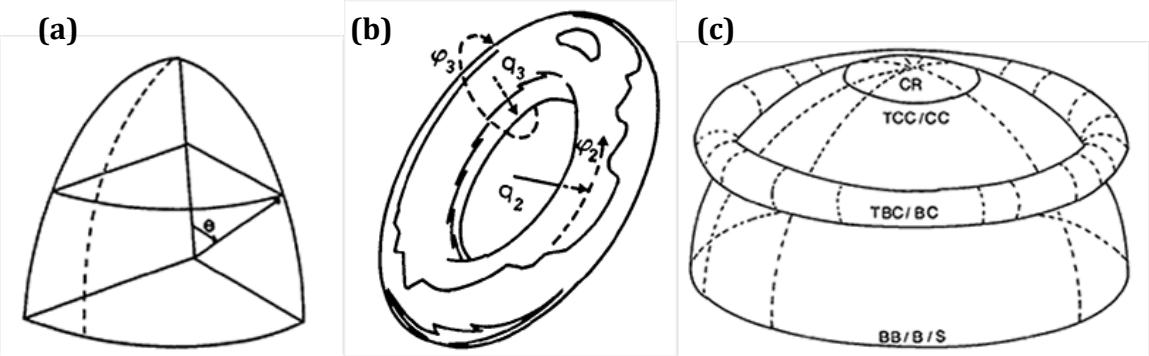


Figure 4.7: Mapping the CP coordinates of an 8-membered ring onto a 3D surface, taken from reference [36]. **(a)** One octant of the unit sphere showing the polar angle θ, **(b)** the coordinates q_2, q_3, Φ_2, and Φ_3 mapped onto a torus, and **(c)** the coordinates θ, q_2, q_3, Φ_2, and Φ_3 mapped onto a 3D surface, by placing the tori at corresponding values of θ.

By taking 2D projections of the 3D surface in **Figure 4.7(c)**, various interconversion paths between the conformations were studied. The CP coordinates of a few homogeneous and heterogeneous 8-membered rings taken from X-ray crystal structures were also calculated and mapped onto the 3D surface [36].

The equivalence properties of CP coordinates have been calculated for medium sized rings [46]. Boeyens and Evans [47] demonstrated that the Cremer and Pople equations (Equations **4.8** to **4.10**) can be mathematically derived from the group theoretical representations of Pickett and Strauss (Equations **4.2** to **4.5**). In this way, it was shown that the PS and CP coordinates are in fact equivalent (summarised in **Table 4.3**).

Table 4.3: Equivalence of PS and CP pucker coordinates for *N*-membered rings.

PS coordinates	CP coordinates
ρ_m	q_m
φ_m	Φ_m
Q	$q_{N/2}$

4.2.3 Intracyclic Torsion Angles

In the next approach, the dihedral angles within the ring (intracyclic torsion angles) are used to derive puckering coordinates. The intracyclic torsion angles are defined *only* by the ring atoms, as depicted schematically in **Figure 4.8**.

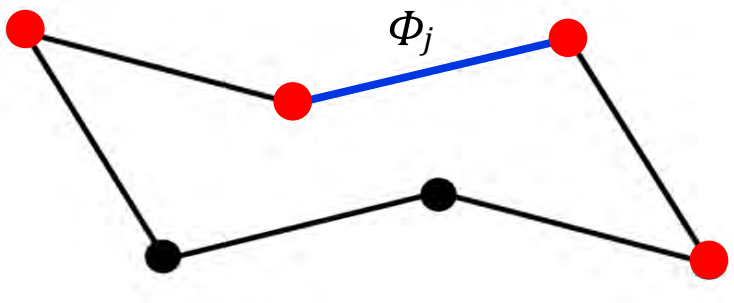
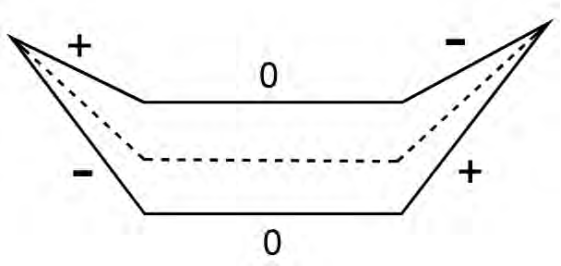


Figure 4.8: Schematic representation of the intracyclic torsion angles Φ_j in a 6-membered ring (the angle Φ_j about the bond in blue is defined by the four red atoms).

In the 1960's, Hendrickson carried out several studies on the canonical conformations of 5- to 10-membered cycloalkanes and methylcycloalkanes, calculating their potential energies and identifying various interconversion paths between the conformers [38, 48-51]. During these studies, he presented a way to characterise the symmetrical ring conformations of 5- to 10-membered cycloalkanes based on the intracyclic torsion angles [38, 49]. Although his method does not involve generating puckering coordinates, it represents a starting point for viewing conformations in terms of their torsion angles. His method is described briefly as follows: the symmetry element (axis or plane of symmetry) passing through the ring is represented as a horizontal line, and the sequence of intracyclic torsion angles about the symmetry element on each side of the ring is indicated by placing +, –, or 0 signs (the signs of the torsion angles) above and below the horizontal line. This system is depicted in **Figure 4.9** for the *boat* conformation of cyclohexane. In **Figure 4.9(a)**, the dashed line is the plane of symmetry passing through two ring atoms, and the signs +, –, and 0 are the signs of the torsion angles about the respective bonds. In **Figure 4.9(b)**, the plane of symmetry is represented as a horizontal line, and the signs of the torsion angles on each side of the plane are written above and below the line.

(a)



(b) $\begin{array}{c} + \ 0 \ - \\ \hline - \ 0 \ + \end{array}$

Figure 4.9: Characterisation of the *boat* conformation of cyclohexane according to Hendrickson [38, 49], adapted from reference [38]. **(a)** The *boat* conformation of cyclohexane, indicating the signs of the intracyclic torsion angles on either side of the plane of symmetry (dashed line). **(b)** The shorthand notation for the boat conformation, where the horizontal line corresponds to the plane of symmetry.

Attempts were then made to generate puckering coordinates in terms of the intracyclic torsion angles. Altona *et al.* [16, 17] related the five intracyclic torsion angles θ_j of a 5-membered ring to the puckering parameters θ_m and P, as given in Equation **4.12** [12].

$$\theta_j = \theta_m \cos(P + 4\pi(j-1)/5) \qquad (\mathbf{4.12})$$

where $1 \leq j \leq 5$ ($j \in \mathbb{Z}$), θ_m is the pucker amplitude and P is the phase angle. This model can be applied to a *general* 5-membered ring (with unequal bond lengths and angles), and is therefore an improvement on the model of Kilpatrick *et al.* (Equation **4.1**) described earlier, which can only be applied to *equilateral* rings. Equation **4.12** is then used to derive Equation **4.13** [12]. The five θ_j angles are used in Equation **4.13** to find the phase P, then θ_m is evaluated using the values of P and θ_1 in Equation **4.12**.

$$\tan P = \frac{\theta_3 + \theta_5 - \theta_2 - \theta_4}{2\,\theta_1[\sin\left(\frac{1}{5}\pi\right) + \sin\left(\frac{2}{5}\pi\right)]} \qquad (\mathbf{4.13})$$

This model has been used to obtain puckering coordinates for several 5-membered rings including a number of purine and pyrimidine glycosides [30, 52, 53]. Díez *et al.* [18] derived a more general expression of Equation **4.12** for rings with large differences in the bond lengths in order to calculate the puckering coordinates for the furanose rings in nucleosides and nucleotides. However, a disadvantage of Equation **4.12** is that it is not an exact relationship, and the value calculated for θ_m is dependent on the numbering of the ring atoms [12]. Equation **4.12** is a good approximation for small magnitudes of puckering [12].

Pucker models for larger rings

Puckering coordinates based on intracyclic torsion angles have also been calculated for several larger ring systems [21, 22, 54-60]. In 1992, Haasnoot [19] described the intracyclic torsion angles ϕ_j of a 6-membered ring as a Fourier series expression, given by Equation **4.14**:

$$\phi_j = \Phi_0 + \Phi_1 \cos\left(P_1 + \frac{2\pi j}{6}\right) + \Phi_2 \cos\left(P_2 + \frac{4\pi j}{6}\right) + \Phi_3 \cos(\pi j) \qquad (\mathbf{4.14})$$

where $0 \leq j \leq 5$ ($j \in \mathbb{Z}$), and the six parameters Φ_0, Φ_1, Φ_2, Φ_3, P_1, and P_2 can be calculated using Fourier inversion equations [19]. The intracyclic torsion angles for experimental structures of a large set of 6-membered ring conformations were then calculated, and using the Fourier inversion equations, it was found that the values of Φ_0 and Φ_1 for these structures were generally very small. Neglecting the terms involving Φ_0 and Φ_1, a simpler expression for ϕ_j is obtained, as given by Equation **4.15**, which is called the Truncated Fourier series model (TF model). Equation **4.15** expresses the torsion angles in terms of just three pucker parameters (i.e. Φ_2, P_2, and Φ_3), instead of the six parameters used in Equation **4.14**. Since certain terms are neglected in Equation **4.15** (as compared to Equation **4.14**), Equation **4.15** is an approximate puckering model for 6-membered rings. Although being an approximate model, it was shown that Equation **4.15** reproduces experimental torsion angles relatively well, using values for Φ_2, P_2, and Φ_3 calculated from Fourier inversion equations) [19].

$$\phi_j = \Phi_2 \cos\left(P_2 + \frac{4\pi j}{6}\right) + \Phi_3 \cos(\pi j) \qquad (\mathbf{4.15})$$

In 2001, Bérces *et al.* [20] used torsion angle vectors to characterise puckers of 6-membered rings. A six-dimensional vector T of the six intracyclic torsion angles τ_i of a general ring conformation was first defined as in Equation **4.16**.

$$T = \{\tau_1, \tau_2, \tau_3, \tau_4, \tau_5, \tau_6\} \qquad (\mathbf{4.16})$$

The torsion vectors for three canonical reference conformations (namely ${}^1C_4$ *chair*, ${}^{1,4}B$ *boat*, and ${}^0S_2$ *twist-boat*) and three 'redundancy' contributions are written out, and the general torsion vector (corresponding to a general conformation) is then expressed as a linear combination of the canonical and redundancy torsion vectors (Equation **4.17**):

$$T = \sum_{i=1}^{3} \lambda_i F_i + \sum_{i=4}^{6} \lambda_i R_i \qquad \textbf{(4.17)}$$

where F_i and R_i correspond to the canonical and redundancy torsion vectors, respectively, and λ_i are the linear combination coefficients. As an example, F_1 is the torsion vector for the $^1C_4$ conformer, shown in Equation **4.18**.

$$F_1 = \{60^\circ, -60^\circ, 60^\circ, -60^\circ, 60^\circ, -60^\circ\} \qquad \textbf{(4.18)}$$

As the ring undergoes out-of-plane motions due to changes in the intracyclic torsion angles (i.e. as it puckers), it can also have some degree of in-plane bending – these two motions are interdependent and give rise to the redundancy terms in Equation **4.17**. The coefficients λ_i are evaluated by projecting the vector T on the six vectors $\{F_i, R_i\}$, where the canonical projection coefficients λ_1, λ_2, and λ_3 measure the puckering distortions of the conformer from the reference ones, while the redundancy projection coefficients λ_4, λ_5, and λ_6 are a measure of the ring bending motions. The larger the coefficient, the more the contribution from the corresponding motion. After characterising some pyranose ring conformations from the literature according to Equation **4.17**, the redundancy coefficients were found to be generally much smaller than the canonical coefficients [20]. Therefore, the ring puckers are essentially defined by the canonical coefficients, which can be viewed as three puckering coordinates for a 6-membered ring.

Attempts were also made to derive pucker coordinates for general *N*-membered rings, such as the model introduced by Zefirov and Palyulin (ZP model) [21, 22]. The ZP model is summarised in **Equations 4.19 – 4.21** [25], and involves a similar mathematical approach as used in the CP model, but instead of using out-of-plane displacements, the ZP model uses a trigonometric function of the intracyclic torsion angles (highlighted in grey) to derive puckering coordinates.

$$s_m \cos\psi_m = -(2/N)^{1/2} \sum_{j=1}^{N} \sin(\phi_j/2) \sin[\pi m(2j+1)/N] \qquad \textbf{(4.19)}$$

$$s_m \sin\psi_m = -(2/N)^{1/2} \sum_{j=1}^{N} \sin(\phi_j/2) \cos[\pi m(2j+1)/N] \qquad \textbf{(4.20)}$$

where $1 \leq j \leq N$ ($j \in \mathbb{Z}$), m = 2, 3, …, (N-1)/2 (when N is odd) and m = 2, 3, ..., ½N-1 (when N is even). When N is even, there is an additional parameter:

$$s_{N/2} = (1/N)^{1/2} \sum_{j=1}^{N} \sin(\phi_j/2) \cos[\pi(j-1)] \qquad \textbf{(4.21)}$$

In the above equations, s_m and ψ_m are the puckering coordinates, and Φ_j is the intracyclic torsion angle between atom numbers j, j+1, j+2, and j+3. The ZP coordinates have been used in the conformational analysis of X-ray structures of several cyclic molecules [22, 25]. Similar to the CP approach where the Cartesian coordinates of the ring atoms can be regenerated from the puckering parameters, the torsion angles in the ZP model can be recalculated from the puckering coordinates, with a small deviation from the original torsion angles (as Equations 4.19 to 4.21 are not mathematically exact) [25].

Puckering models based on torsion angles are particularly advantageous for studying the conformations of cyclic molecules in solution, since experimentally-determined NMR coupling constants can be used to calculate ring torsion angles, which can then be used to obtain puckering coordinates. These puckering models have in the past been advantageous over models based on Cartesian coordinates – for example, according to Haasnoot when using the CP model, the Cartesian coordinates of the ring atoms must be known, which would be impossible to determine for molecules in solution [19]. However, Cremer has since shown that is it possible to calculate NMR spin-spin coupling constants using the CP coordinates [61, 62]. CP coordinates can also be used in connection with internal and curvilinear coordinates [63, 64].

4.2.4 Triangular Tessellation

This method has previously been introduced as a decomposition method but it can instead be seen as a 'voronoi-like' tessellation of the ring where only triangular pieces are allowed. The angles between these pieces relative to the centre triangle is used as a means to measure puckering conformation. In this approach, the ring is tessellated into triangular planes and a ***new*** set of internal coordinates based on angular rotations of these planes is defined. These internal coordinates are used directly as the puckering coordinates, and will be called ***Angular Puckering Coordinates*** **(APC)** herein. The method to generate APC will be called the ***Triangular Tessellation*** **(TT)** method. The

term *Triangular Decomposition* was first used by Hill and Reilly in 2007 [10] to describe this method, but the method was in fact introduced much earlier in the 1970's.

In 1970, Pickett and Strauss [23] introduced the triangular tessellation method to analyse the conformations of 6-membered rings such as cyclohexane, p-dioxane and tetrahydropyran. As depicted in **Figure 4.10**, the 6-membered ring is broken down into four triangular planes by drawing in axes between three alternating ring atoms. The central plane is viewed as the reference plane, and the remaining three planes are treated as rotatable planes, or colloquially, 'flaps' that can rotate relative to the reference plane about the respective axes – the rotations are measured quantitatively by a newly-defined set of internal coordinates (or APC) α_i. Since the α_i rotations are out-of-plane bending modes, they can be used to describe the puckering of the ring, because. as α_i changes, the ring takes on different conformations.

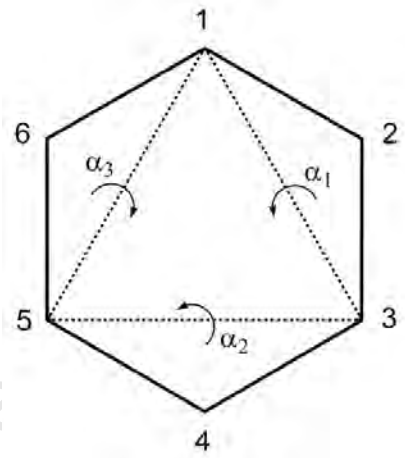


Figure 4.10: Triangular tessellation of a 6-membered ring, adapted from reference [23]. The APC α_1, α_2, α_3 represent out-of-plane rotations about the axes.

In their work, Pickett and Strauss do not use α_i as pucker coordinates, but rather express α_i in terms of the variables r, θ, and Φ, as given by Equation **4.22**:

$$\alpha_i = r\left\{\cos\theta + 2\left[\cos\left(\frac{4\pi i}{3} - \frac{2\pi}{3} + \phi\right)\right]\sin\theta\right\} \qquad (\mathbf{4.22})$$

where i = 1, 2, 3. The variables r, θ, and Φ are then used as the pucker coordinates, where r measures the ring's deviation from planarity, and the angles θ and Φ represent the actual shape of the molecule. Using these puckering coordinates together with a potential energy function, conformational energy maps of the molecules are generated [23] – the map of cyclohexane was used the following year by Strauss [65] to look at the

process of inversion between the *chair* forms of cyclohexane. Then in 1979, using the same approach as reference [23], Joshi and Rao [24] investigated the flexibility of the conformations of α- and β-D-glucopyranose by generating and analysing their conformational energy maps.

For 7-membered rings, Bocian *et al.* [14] presented a set of four APC to characterise the conformations of cycloheptane. These APC are labelled α_1, α_2, α_3, and δ (**Figure 4.11**), and are analagous to those defined previously by Pickett and Strauss for 6-membered rings [23]. In this case, two reference planes are needed to fully describe the puckering motions of the ring. As depicted in **Figure 4.11**, the plane defined by atoms 1, 3, and 5 serves as a reference plane for the α_1, α_2, and δ rotations, while the plane responsible for the δ rotation also serves as a reference plane for the α_3 rotation.

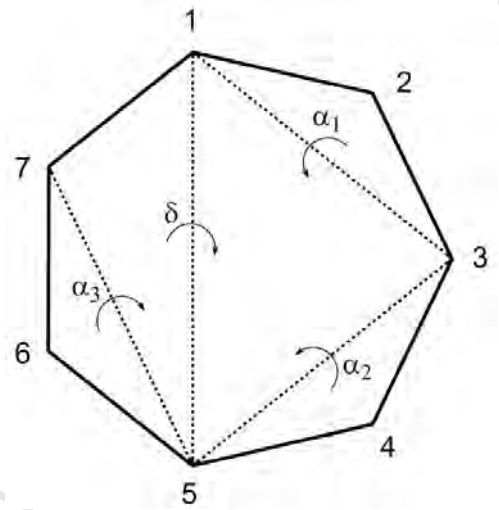


Figure 4.11: Triangular tessellation of a 7-membered ring, adapted from reference [14]. The APC α_1, α_2, α_3 and δ represent rotations about the axes.

In 2007, Hill and Reilly [10] studied the triangular tessellation of cyclohexane, using the same system of APC developed previously [23]. In addition, they presented a detailed mathematical procedure to calculate the APC of 6-membered rings. Using their procedure, Hill and Reilly generated APC for the 38 canonical conformations of cyclohexane, which is tabulated in their paper [10] – their procedure will be discussed in more detail in Chapter 5.

A recent application of puckering coordinates based on triangular tessellation has been presented by Barnett and Naidoo [5], where the FEARCF method was used to plot the

free energy surface of the conformations of β-D-glucose in terms of the three APC for 6-membered rings. The APC were plotted in three dimensions, while the relative free energies were plotted in the fourth dimension using a colouring scheme to indicate low and high energy conformers, as depicted in **Figure 4.12**.

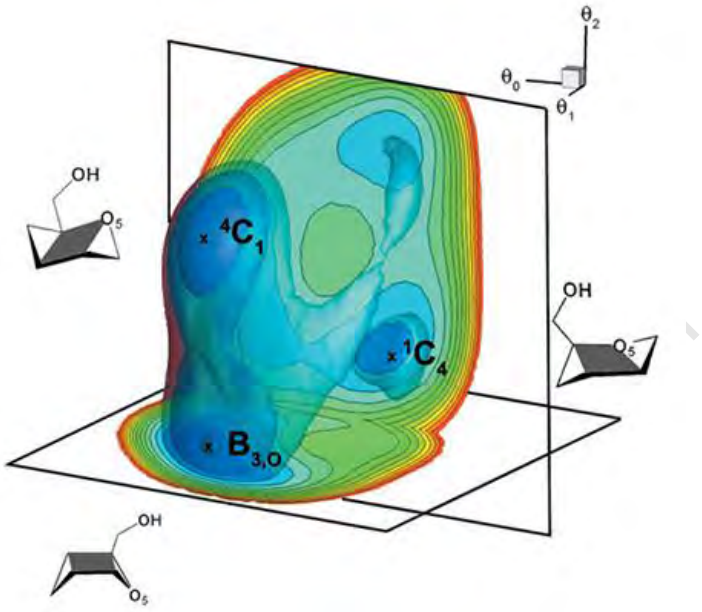


Figure 4.12: The 3D free energy surface of β-D-glucose (from reference [5]), plotted in terms of the angular puckering coordinates for 6-membered rings, (θ_0, θ_1, θ_2). The location of three conformers has been indicated. High and low energy conformations are indicated in red and blue colour, respectively.

There are both advantages and limitations to the triangular tessellation method. One advantage is that one can easily picture the triangular planes of the puckered ring as tilted above or below the corresponding reference plane, which makes APC a simple way to visualise ring conformations for small rings (for 5 and 6-membered rings, the sign and magnitude of the APC are directly related to a particular geometrical element of the ring). In the case of 6-membered rings for example, the $^4C_1$ *chair* conformation of cyclohexane has APC of (35.26°, 35.26°, 35.26°) [10], where the positive values indicate that the triangular planes lie above the reference plane, while the $^1C_4$ *chair* conformation has APC of (-35.26°, -35.26°, -35.26°) [10], with the negative values indicating the triangular planes to lie below the reference plane.

A limitation of APC is that the same ring conformational shape can have slightly different values for the APC, depending on the extent to which the ring is puckered. For

example, a molecular dynamics trajectory of cyclohexane could generate $^4C_1$ *chair* conformations with APC of (35.26°, 35.26°, 35.26°) as well as (32°, 32°, 32°), where the former corresponds to the ideal conformation [10] while the latter represents the same conformation with a lesser extent of pucker. This can be a complicating factor when attempting to assign the pucker conformations generated during a computer simulation.

A further limitation of this method is that it is not a unique measurement. There are several ways of tessellating a ring and a particular 'root atom' and direction of measurement (which is the reference plane, which direction is up and which direction is clockwise?) must be chosen when comparing puckers across several chemical systems. This problem is not limited to the tessellation approach. The frame of reference for defining ring pucker is a key choice when using pucker methods.

Using the TT model in this work

In this work, the TT model is used to quantify the macrocyclic puckering of α-, β- and γ-CD in water (Chapter 6). The reasons for choosing the TT model have been explained in Chapter 1, and are mentioned briefly as follows:

Although APC based on the TT model are usually used to describe monocyclic ring conformations, they can also be applied to the *macrocyclic* CD rings in this thesis through a suitable coarse graining of the CDs. Furthermore, since the APC are all the same type of quantity (i.e. angular quantities) and the conformational motions across the *homogeneous* CD rings are measured over a long simulation time, *average* time correlation functions can be calculated for the APC of each CD – these correlation functions can then be compared for the three CDs to gain insight into their relative puckering flexibilities.

References

1. Dobler, M.; Dunitz, J. D.; Seiler, P. *Acta Crystallographica Section B* **1974**, *30*, 2741.
2. Lindoy, L. F. *The Chemistry of Macrocyclic Ligand Complexes*; Cambridge University Press: Cambridge, 1989.
3. Naidoo, K. J.; Gamieldien, M. R.; Chen, J. Y.-J.; Widmalm, G. r.; Maliniak, A. *The Journal of Physical Chemistry B* **2008**, *112*, 15151.
4. Seiler, P.; Dobler, M.; Dunitz, J. D. *Acta Crystallographica Section B* **1974**, *30*, 2744.
5. Barnett, C. B.; Naidoo, K. J. *Molecular Physics* **2009**, *107*, 1243.
6. Biarnés, X.; Ardèvol, A.; Planas, A.; Rovira, C.; Laio, A.; Parrinello, M. *Journal of the American Chemical Society* **2007**, *129*, 10686.
7. Brayer, G. D.; Sidhu, G.; Maurus, R.; Rydberg, E. H.; Braun, C.; Wang, Y.; Nguyen, N. T.; Overall, C. M.; Withers, S. G. *Biochemistry* **2000**, *39*, 4778.
8. Gloster, T. M.; Davies, G. J. *Organic & Biomolecular Chemistry* **2010**, *8*, 305.
9. Schramm, V. L. *Annual Review of Biochemistry* **1998**, *67*, 693.
10. Hill, A. D.; Reilly, P. J. *Journal of Chemical Information and Modeling* **2007**, *47*, 1031.
11. Kilpatrick, J. E.; Pitzer, K. S.; Spitzer, R. *J Am Chem Soc* **1947**, *69*, 2483.
12. Cremer, D.; Pople, J. A. *Journal of the American Chemical Society* **1975**, *97*, 1354.
13. Pickett, H. M.; Strauss, H. L. *The Journal of Chemical Physics* **1971**, *55*, 324.
14. Bocian, D. F.; Pickett, H. M.; Rounds, T. C.; Strauss, H. L. *J Am Chem Soc* **1975**, *97*, 687.
15. Evans, D. G.; Boeyens, J. C. A. *Acta Crystallographica Section B* **1989**, *45*, 581.
16. Altona, C.; Geise, H. J.; Romers, C. *Tetrahedron* **1968**, *24*, 13.
17. Geise, H. J.; Altona, C.; Romers, C. *Tetrahedron Letters* **1967**, *8*, 1383.
18. Diez, E.; Esteban, A. L.; Bermejo, F. J.; Rico, M. *The Journal of Physical Chemistry* **1980**, *84*, 3191.
19. Haasnoot, C. A. G. *J Am Chem Soc* **1992**, *114*, 882.
20. Bérces, A.; Whitfield, D. M.; Nukada, T. *Tetrahedron* **2001**, *57*, 477.
21. Zefirov, N. S.; Palyulin, V. A. *Dokl. Akad. Nauk SSSR.* **1980**, *252*, 111.
22. Zefirov, N. S.; Palyulin, V. A.; Dashevskaya, E. E. *Journal of Physical Organic Chemistry* **1990**, *3*, 147.
23. Strauss, H. L.; Pickett, H. M. *J Am Chem Soc* **1970**, *92*, 7281.

24. Joshi, N. V.; Rao, V. S. R. *Biopolymers* **1979**, *18*, 2993.
25. Zotov, A. Y.; Palyulin, V. A.; Zefirov, N. S. *Journal of Chemical Information and Computer Sciences* **1997**, *37*, 766.
26. International Union of Pure and Applied Chemistry and International Union of Biochemistry Joint Commission on Biochemical Nomenclature (JCBN) *European Journal of Biochemistry* **1980**, *111*, 295.
27. International Union of Pure and Applied Chemistry and International Union of Biochemistry and Molecular Biology Joint Commission on Biochemical Nomenclature (JCBN) *Pure and applied chemistry* **1996**, *68*, 1919.
28. Schwarz, J. C. P. *Journal of the Chemical Society, Chemical Communications* **1973**, 505.
29. Stoddart, J. F. *Stereochemistry of carbohydrates*; John Wiley & Sons: New York, 1971.
30. Altona, C.; Sundaralingam, M. *J Am Chem Soc* **1972**, *94*, 8205.
31. Ionescu, A. R.; Bérces, A.; Zgierski, M. Z.; Whitfield, D. M.; Nukada, T. *The Journal of Physical Chemistry A* **2005**, *109*, 8096.
32. Boeyens, J. C. A. *Journal of Crystal and Molecular Structure* **1978**, *8*, 317.
33. DeMatteo, M. P.; Snyder, N. L.; Morton, M.; Baldisseri, D. M.; Hadad, C. M.; Peczuh, M. W. *Journal of Organic Chemistry* **2005**, *70*, 24.
34. Wiberg, K. B. *Journal of Organic Chemistry* **2003**, *68*, 9322.
35. Boessenkool, I. K.; Boeyens, J. C. A. *Journal of Crystal and Molecular Structure* **1980**, *10*, 11.
36. Evans, D. G.; Boeyens, J. C. A. *Acta Crystallographica Section B-Structural Science* **1988**, *44*, 663.
37. Kleinpeter, E. *Advances in Heterocyclic Chemistry* **2004**, *86*, 41.
38. Hendrickson, J. B. *Journal of the American Chemical Society* **1967**, *89*, 7047.
39. Geise, H. J.; Adams, W. J.; Bartell, L. S. *Tetrahedron* **1969**, *25*, 3045.
40. Herzyk, P.; Rabczenko, A. *Journal of the Chemical Society, Perkin Transactions 2* **1985**, 1925.
41. Adams, W. J.; Geise, H. J.; Bartell, L. S. *J Am Chem Soc* **1970**, *92*, 5013.
42. Herzyk, P.; Rabczenko, A. *Journal of the Chemical Society, Perkin Transactions 2* **1983**, 213.

43. Cremer, D.; Szabó, K. J. In *Conformational Behavior of Six-Membered Rings: Analysis, Dynamics, and Stereoelectronic Effects*; Juaristi, E., Ed.; VCH Publishers: 1995, p 59.
44. Autieri, E.; Sega, M.; Pederiva, F.; Guella, G. *The Journal of Chemical Physics* **2010**, *133*, 095104.
45. Sega, M.; Autieri, E.; Pederiva, F. *The Journal of Chemical Physics* **2009**, *130*, 225102.
46. Kessler, M.; Pérez, J. *J Math Chem* **2012**, *50*, 187.
47. Evans, D. G.; Boeyens, J. C. A. *Acta Crystallographica Section B* **1989**, *45*, 577.
48. Hendrickson, J. B. *Journal of the American Chemical Society* **1961**, *83*, 4537.
49. Hendrickson, J. B. *Journal of the American Chemical Society* **1964**, *86*, 4854.
50. Hendrickson, J. B. *Journal of the American Chemical Society* **1967**, *89*, 7036.
51. Hendrickson, J. B. *Journal of the American Chemical Society* **1967**, *89*, 7043.
52. Altona, C.; van der Veek, A. P. M. *Tetrahedron* **1968**, *24*, 4377.
53. Romers, C.; Altona, C.; Buys, H. R.; Havinga, E. *Top. Stereochem.* **1969**, *4*, 39.
54. Buys, H. R.; Geise, H. J. *Tetrahedron Letters* **1968**, 5619.
55. cano, F. H.; Foces-Foces, C.; García-Blanco, S. *Tetrahedron* **1977**, *33*, 797.
56. Cano, F. H.; Focesfoces, C.; Garciablanco, S. *Acta Crystallographica Section A* **1978**, *34*, S91.
57. Díez, E.; Esteban, A. L.; Guilleme, J.; Bermejo, F. L. *Journal of Molecular Structure* **1981**, *70*, 61.
58. Díez, E.; Esteban, A. L.; Bermejo, F. J.; Altona, C.; de Leeuw, F. A. A. M. *Journal of Molecular Structure* **1984**, *125*, 49.
59. de Leeuw, F. A. A. M.; Van Kampen, P. N.; Altona, C.; Díez, E.; Esteban, A. L. *Journal of Molecular Structure* **1984**, *125*, 67.
60. Cano, F. H.; Foces-Foces, C. *Journal of Molecular Structure* **1983**, *94*, 209.
61. Wu, A.; Cremer, D.; Auer, A. A.; Gauss, J. *The Journal of Physical Chemistry A* **2002**, *106*, 657.
62. Wu, A.; Cremer, D. *The Journal of Physical Chemistry A* **2003**, *107*, 1797.
63. Essen, H.; Cremer, D. *Acta Crystallographica Section B* **1984**, *40*, 418.
64. Cremer, D. *The Journal of Physical Chemistry* **1990**, *94*, 5502.
65. Strauss, H. L. *Journal of Chemical Education* **1971**, *48*, 221.

Chapter 5*

Defining Pucker Coordinates for 6-, 7- and 8-membered Rings

Cyclic conformational coordinates are essential for the distinction of molecular ring conformers as the use of Cremer-Pople coordinates have illustrated for 5- and 6-membered rings. In this chapter, the triangular tessellation (TT) approach [1] is extended, by tessellating medium rings into triangular planes and using the relative angles made between triangular planes. Triangular tessellation and assignment of 7- and 8-membered rings is illustrated using sets of canonical states for cycloheptane and cyclooctane that have been previously experimentally analysed. A systematic procedure to generate the angular puckering coordinates (APC) of 6-, 7- and 8-membered rings by triangular tessellation of the rings is also developed.

5.1 Canonical Conformations of 6-, 7- and 8-membered Rings

Small (nodes (N) $\leq$ 6) and medium ($7 \leq N \leq 13$) sized molecular rings display a preference for definite canonical conformations that are linked to their physical and chemical properties. For instance, the canonical shapes of a 6-membered ring, the *chair* (*C*), *boat* (*B*), *half-chair* (*H*), *skew-boat* (*S*), and *envelope* (*E*) conformations [2] were discussed in Chapter 4 and depicted in **Figure 4.1**.

The puckers of 7- and 8-membered rings can also be classified according to a system of canonical shapes, as has been done in conformational studies of cycloheptane [3-5] and cyclooctane [5-7], as well as other 7- and 8-membered rings [8, 9]. These studies involved locating minimum energy structures and transition states on the potential energy surface of cyclooctane [7]; studying the relative energies [5], geometries [6], and pseudorotational pathways (conformational paths) [4] of cycloheptane and cyclooctane conformers; calculating pucker parameters of eight-membered ring conformations from their crystal structures [9]; characterising the preferred conformations of two

* This chapter is based on a published paper **"Interpreting medium ring canonical conformers by a triangular plane tessellation of the macrocycle"** J. Chem. Phys. **138**, 184110 (2013); http://dx.doi.org/10.1063/1.4803698

septanose carbohydrates in solution [8]; and identifying the lowest energy conformers of cycloheptane and selected oxepanes [3].

In the study carried out by Wiberg [5] in 2003, Wiberg carried out geometry optimisations on the five canonical shapes of cycloheptane and eleven shapes of cyclooctane at the MP2/6-311+G* level of theory, listing the Cartesian coordinates of the optimised structures in his supporting information [5]. The relative energies of the conformers predicted by the MP2 level of theory appear to give good agreement with experimental data [5], and in this thesis, the shapes presented by Wiberg will be considered. These shapes are tabulated in **Tables 5.1** and **5.2**, and illustrated in **Figures 5.1** and **5.2**, for cycloheptane and cyclooctane, respectively.

Table 5.1: The five canonical shapes for cycloheptane, as studied by Wiberg [5].

Canonical shape	Symbol
Twist-chair	TC
Twist-boat	TB
Chair	C
Boat	B
T_3	T_3

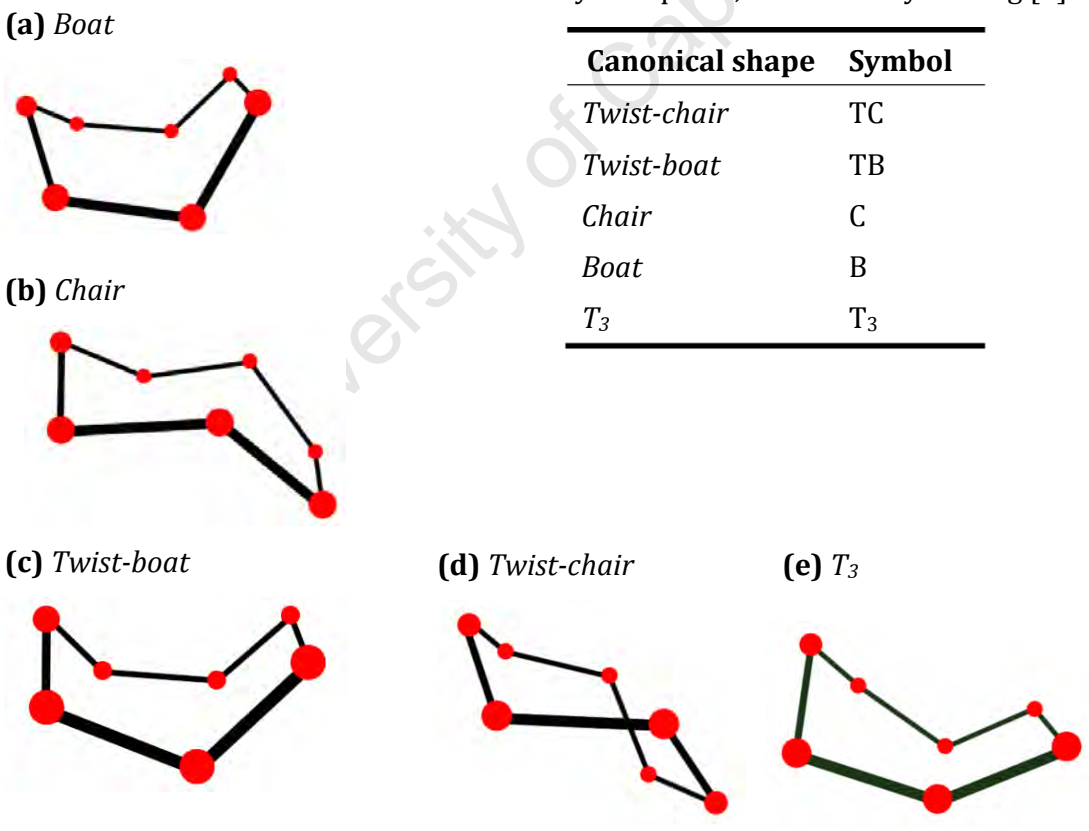


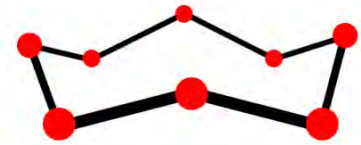
Figure 5.1: The five canonical shapes of cycloheptane, redrawn from the Cartesian coordinates given by Wiberg [5]. T_3 is a transition state connecting the *twist-chair* and *twist-boat* conformations.

Table 5.2: The eleven canonical shapes of cyclooctane, as studied by Wiberg [5].

Canonical shape	Symbol
Boat-chair	BC
Twist boat-chair	TBC
Boat	B
Chair	C
Crown	Cr
Boat-boat	BB
Twist chair-chair	TCC
TS1	TS1
TS2	TS2
TS3	TS3
TS4	TS4

(a) *Chair*

(b) *Crown*



(c) *Boat-chair*

(f) *Twist boat-chair*

(i) *TS2*

(d) *Boat*

(g) *Twist chair-chair*

(j) *TS3*

(e) *Boat-boat*

(h) *TS1*

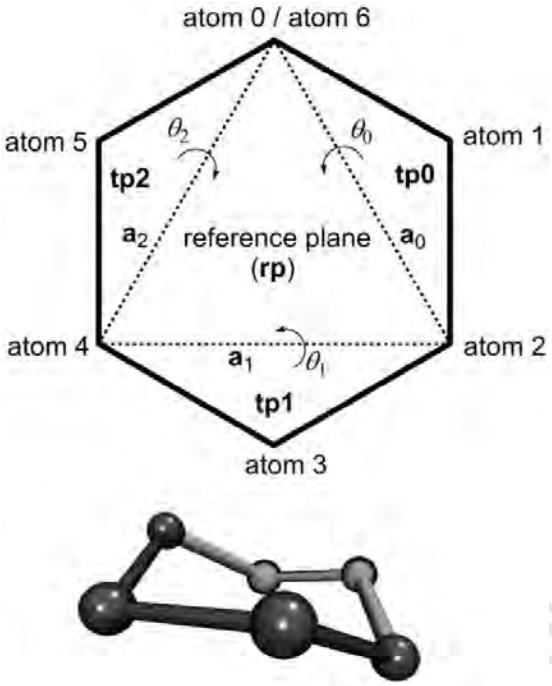
(k) *TS4*

Figure 5.2:. The eleven canonical shapes of cyclooctane, redrawn from the Cartesian coordinates given by Wiberg [5]. The various transition state structures are labelled *TSi*, where *i* = 1, 2, 3, 4.

5.2 Triangular Tessellation of 6-membered Rings

The method of tessellating a macrocyclic ring into triangular planes was introduced in the 1970's for 6-membered [10, 11] and 7-membered rings [12]. The method did not draw much interest and attention until recently when Hill and Reilly [1] illustrated it on six membered rings while coining the term *triangular decomposition* – however, in this thesis, the term *triangular tessellation* will be used. The triangular tessellation of a 6-membered ring is depicted in **Figure 5.3**, and the details of this method are outlined below as it facilitates the introduction of this approach to medium sized rings.

The ring atoms are numbered atom i in a clockwise manner, where $0 \leq i \leq 6$ and atom 0 ≡ atom 6 – this will be referred to as ***atom-numbering*** here (it is ***not*** related to the IUPAC numbering system mentioned in Chapter 4). Three alternating ring atoms (atom 0, atom 2, atom 4) are chosen to define a reference plane (**rp**), and the remaining three triangular planes **tp*i*** (***i*** = 0, 1, 2), colloquially referred to as 'triangular flaps' are defined by the sets of atoms (atom 0, atom 1, atom 2), (atom 2, atom 3, atom 4), and (atom 4, atom 5, atom 6), respectively. The **tp**s rotate relative to **rp** about axes of puckering $\mathbf{a}_i$, where $0 \leq i \leq 2$ – the axes of puckering are shown as dotted lines in **Figure 5.3**. The angles between the **tp**s and **rp** are labelled as θ_i, where $0 \leq i \leq 2$ and $-90° \leq \theta_i \leq 90°$. For example, **tp0** rotates relative to **rp** about the axis $\mathbf{a}_0$ and makes the angle θ_0 with **rp**. Each conformation of the ring can be described by a unique coordinate set (θ_0, θ_1, θ_2) – these angles quantitatively measure the puckering of the ring, and are collectively referred to as angular puckering coordinates (APC). Since $-90° \leq \theta_i \leq 90°$, the ring conformers lie in a bounded region of TT pucker coordinate space. The *chair* conformation for the 6- membered ring along with a set of APC is shown in **Figure 5.3.**



CONFORMER	ANGULAR PUCKERING COORDINATES
Chair	(-35.0°, -35.0°, -35.0°)

Figure 5.3: A tessellation of a 6-membered ring into triangular planes. Below this, the chair conformation is shown with the associated angles that the **tp**s make with the **rp**.

The derivation of a mathematical expression for the APC is taken from reference [1] and is described below and summarised in Equations **5.1a** to **5.1g** [1], where vector quantities are written in bold script. The Cartesian coordinates, $\mathbf{x}_i$, of the ring atoms, are used to define the axes of puckering, $\mathbf{a}_i$, as shown in Equations **5.1a** and **5.1b**. In Equation **5.1a**, $\mathbf{x}_0 \equiv \mathbf{x}_6$.

$$\mathbf{x}_i = \langle \mathrm{x}_i, \mathrm{y}_i, \mathrm{z}_i \rangle \qquad \text{for } 0 \leq i \leq 6 \qquad \textbf{(5.1a)}$$

$$\mathbf{a}_i = \mathbf{x}_{2(i+1)} - \mathbf{x}_{2i} \qquad \text{for } 0 \leq i \leq 2 \qquad \textbf{(5.1b)}$$

A vector normal to **rp**, **n**, is calculated by taking the cross product of two axes of puckering (Equation **5.1c**), where $\otimes$ denotes the vector cross product.

$$\mathbf{n} = \mathbf{a}_1 \otimes \mathbf{a}_0 \qquad \textbf{(5.1c)}$$

Bond vectors, $\mathbf{r}_i$, lying along the bonds between the ring atoms are defined in Equation **5.1d**. The bond vectors are then used to define the orientation vectors, $\mathbf{p}_i$, of the central atom of each **tp** [Equation **5.1e**).

$$\mathbf{r}_i = \mathbf{x}_{i+1} - \mathbf{x}_i \qquad \text{for } 0 \leq i \leq 5 \qquad \textbf{(5.1d)}$$

$$\mathbf{p}_i = \mathbf{r}_{i-1} \otimes \mathbf{r}_i \qquad \text{for } i \in \{1, 3, 5\} \qquad \textbf{(5.1e)}$$

The axes of puckering and orientation vectors are used to define vectors $\mathbf{q}_i$, which lie along the perpendicular (⊥) line joining $\mathbf{a}_i$ and $\mathbf{x}_{i+1}$ (Equation **5.1f**).

$$\mathbf{q}_i = \mathbf{a}_i \otimes \mathbf{p}_{2i+1} \qquad \text{for } 0 \le i \le 2 \qquad \textbf{(5.1f)}$$

The pucker angle each **tp** makes with **rp**, θ_i, is then calculated by Equation **5.1g**.

$$\theta_i = \pi/2 - \cos^{-1}\left[\frac{\mathbf{q}_i \odot \mathbf{n}}{|\mathbf{q}_i| \times |\mathbf{n}|}\right] \qquad \text{for } 0 \le i \le 2, \text{ where } -90^{o} \le \theta_i \le 90^{o} \qquad \textbf{(5.1g)}$$

As the **tp**s rotate relative to **rp**, the vectors $\mathbf{p}_i$ and $\mathbf{q}_i$ change direction and the angles θ_i change magnitude and sign. Triangular planes positioned above and below the reference plane have positive and negative APC, respectively [1]. **Figure 5.4** is an illustration of the triangular tessellation of a 6-membered ring (adapted from **Figure 2** in reference [1]), showing the various vectors, **tp**s, and pucker angles. The axes are shown as dotted lines in **Figure 5.4**, but they are in fact vector quantities.

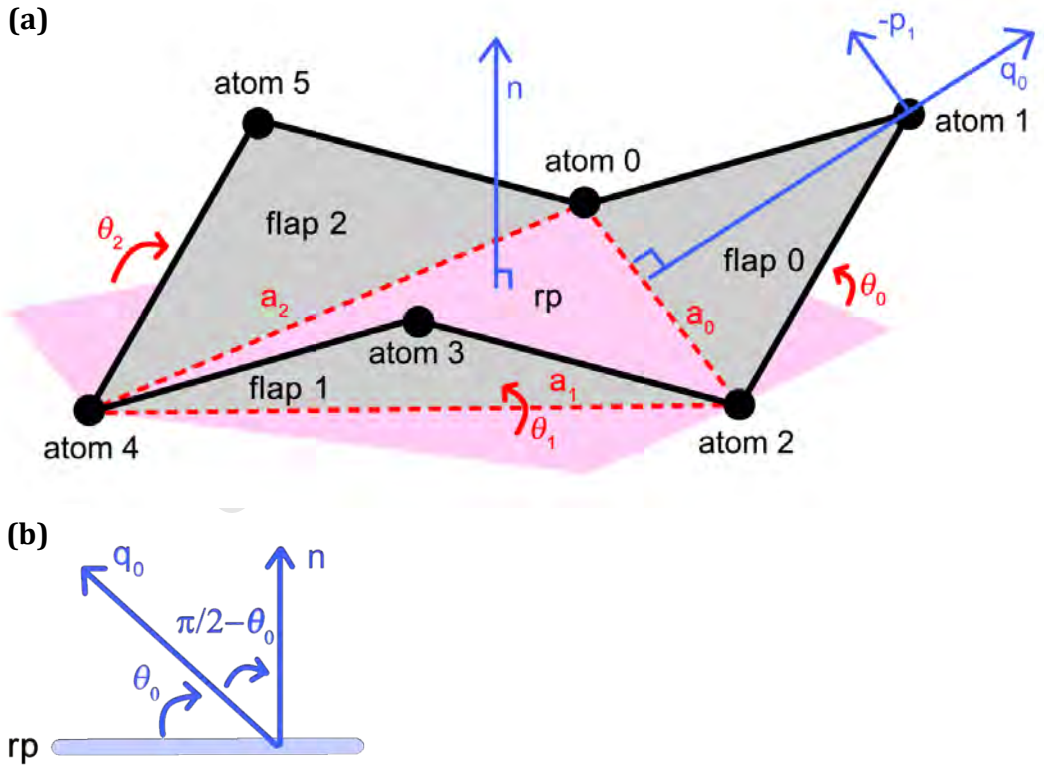
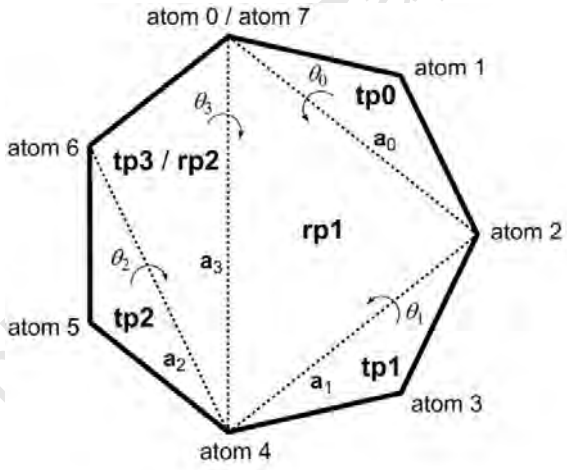


Figure 5.4: (a) Representation of the triangular tessellation of the *chair* shape of a 6-membered ring, adapted from reference [1]. The directions of vectors $\mathbf{q}_0$ and **n** drawn here are consistent with using the right-hand rule (these vectors are drawn in opposite directions in reference [1]). The reference plane is extended outwards to show the APC more easily and the vectors are not drawn to scale. **(b)** Diagram showing the angles made between vectors $\mathbf{q}_0$ and **n**, and between $\mathbf{q}_0$ and the reference plane.

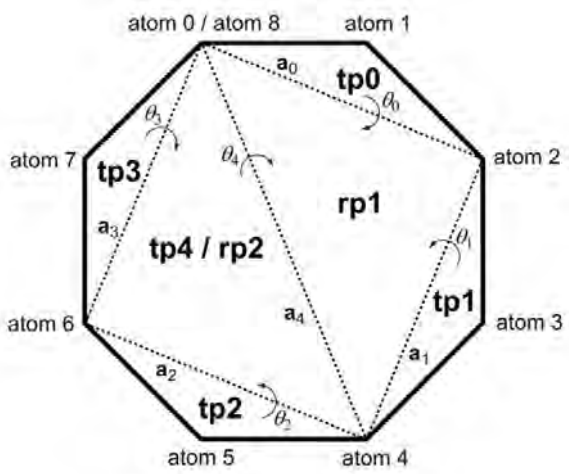
5.3 Triangular Tessellation of 7- and 8-membered Rings

In this work, the TT approach, which had been used to analyse small 6-membered rings, is extended to medium sized odd numbered 7-membered and even numbered 8-membered rings. Triangular Tessellation of 7- and 8-membered rings is depicted in **Figures 5.5** and **5.6**, respectively. In each case, the ring atoms are numbered atom *i* in a clockwise manner, where $0 \leq i \leq N$ and atom $0 \equiv$ atom N (N is the number of ring atoms). Compared with small rings, such as the 6-membered ring described above, more than one **rp** is required for medium-sized rings. Particularly for 7- and 8-membered rings, two **rp**s must be specified to unambiguously define the positions of all **tp**s. For example, in **Figure 5.5** (which is adapted from **Figure 3** of reference [1]), **tp0** rotates relative to reference plane 1 (**rp1**) while **tp2** rotates relative to reference plane 2 (**rp2**). However, the reference planes may rotate relative to each other. Therefore when **rp2** rotates relative to **rp1**, it can be referred to as **tp3** as it plays the role of both reference and rotating triangular plane.



CONFORMER	ANGULAR PUCKERING COORDINATES
Chair	(55.8°, 27.6°, 55.8°, 0.0°)

Figure 5.5: A tessellation of a 7-membered ring into triangular planes (adapted from **Figure 3** of reference [1]) and an illustration of a *chair* conformation and the associated angles that the **tp**s make with the **rp**s.



CONFORMER	ANGULAR PUCKERING COORDINATES
Chair	(79.7°, 31.9°, -79.7°, -31.9°, 0.0°)

Figure 5.6: A tessellation of an 8-membered ring into triangular planes. Below it is a *chair* conformation and the associated angles that the **tp**s make with the **rp**s.

Both the 7- and 8-membered rings are decomposed into two reference planes (**rp*i***, ***i*** = 1, 2) and N-3 triangular planes (i.e. **tp*i***, ***i*** = 0, 1, 2, … , N-4). Thus in **Figure 5.5**, **rp2** also serves as **tp3**, and in **Figure 5.6**, **rp2** also serves as **tp4**. The axes of puckering and the pucker angles are labelled $\mathbf{a}_i$ and θ_i, respectively, where $0 \leq i \leq N\text{-}4$ with $-90° \leq \theta_i \leq 90°$. For 7-membered rings, there are four APC for each conformation, denoted by the set $(\theta_0, \theta_1, \theta_2, \theta_3)$, while for 8-membered rings, there are five APC, denoted by $(\theta_0, \theta_1, \theta_2, \theta_3, \theta_4)$, i.e. since there are always N-3 APC for an N-membered ring, the 6-, 7- and 8-membered rings have three, four and five APC, respectively.

The Cartesian coordinates of the ring atoms are used to derive expressions to compute the APC made by the tessellated **tp**s of a 7-membered ring. The derivation for 7-membered rings is outlined Equations **5.2a – l**. Axes of puckering, $\mathbf{a}_i$, are defined based on the Cartesian coordinates of the ring atoms, $\mathbf{x}_i$, as given by Equations **5.2a – c**. In Equation **5.2a**, $\mathbf{x}_0 \equiv \mathbf{x}_7$.

$$\mathbf{x}_i = \langle x_i, y_i, z_i \rangle \qquad \text{for } 0 \le i \le 7 \qquad \textbf{(5.2a)}$$

$$\mathbf{a}_i = \mathbf{x}_{2(i+1)} - \mathbf{x}_{2i} \qquad \text{for } 0 \le i \le 2 \qquad \textbf{(5.2b)}$$

$$\mathbf{a}_3 = \mathbf{x}_0 - \mathbf{x}_4 \qquad \textbf{(5.2c)}$$

Vectors normal to **rp1** and **rp2**, namely $\mathbf{n}_1$ and $\mathbf{n}_2$ respectively, are calculated as the cross product of two axes of puckering, as shown in Equations **5.2d** – **e**.

$$\mathbf{n}_1 = \mathbf{a}_1 \otimes \mathbf{a}_0 \qquad \textbf{(5.2d)}$$

$$\mathbf{n}_2 = \mathbf{a}_3 \otimes \mathbf{a}_2 \qquad \textbf{(5.2e)}$$

Bond vectors, $\mathbf{r}_i$, are defined and used to define orientation vectors, $\mathbf{p}_i$, for **tp**s 0 to 3, given in Equations **5.2f** – **h**. The orientation vector for **tp3** is denoted by $\mathbf{p}_{\mathbf{tp3}}$, and is given in Equation **5.2h**.

$$\mathbf{r}_i = \mathbf{x}_{i+1} - \mathbf{x}_i \qquad \text{for } 0 \le i \le 6 \qquad \textbf{(5.2f)}$$

$$\mathbf{p}_i = \mathbf{r}_{i-1} \otimes \mathbf{r}_i \qquad \text{for } i \in \{1, 3, 5\} \qquad \textbf{(5.2g)}$$

$$\mathbf{p}_{\mathbf{tp3}} = \mathbf{a}_2 \otimes \mathbf{r}_6 \qquad \textbf{(5.2h)}$$

The axes of puckering and orientation vectors are used to define the vectors $\mathbf{q}_i$ for **tp**s 0 to 2, given by Equation **5.2i**; $\mathbf{q}_{\mathbf{tp3}}$ (for **tp3** of the ring) is a special case, defined separately in Equation **5.2j**. $\mathbf{q}_{\mathbf{tp3}}$ (also called $\mathbf{q}_3$) lies along the ⊥ line joining $\mathbf{a}_3$ and $\mathbf{x}_6$, depicted later in **Figure 5.7** (where $\mathbf{x}_6$ corresponds to atom 6).

$$\mathbf{q}_i = \mathbf{a}_i \otimes \mathbf{p}_{2i+1} \qquad \text{for } 0 \le i \le 2 \qquad \textbf{(5.2i)}$$

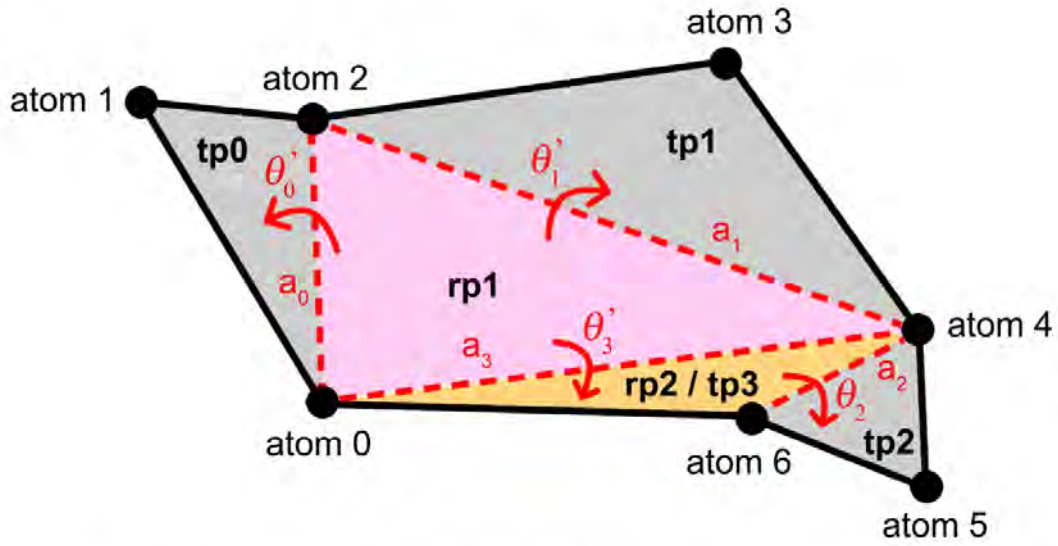
$$\mathbf{q}_{\mathbf{tp3}} = \mathbf{q}_3 = \mathbf{a}_3 \otimes \mathbf{p}_{\mathbf{tp3}} \qquad \textbf{(5.2j)}$$

The pucker angles are then calculated in Equations **5.2k – l**.

$$\theta_i = \pi/2 - \cos^{-1}\left[\frac{\mathbf{q}_i \odot \mathbf{n}_1}{|\mathbf{q}_i| \times |\mathbf{n}_1|}\right] \qquad \text{for } i \in \{0, 1, 3\}, \text{ where } -90^o \le \theta_i \le 90^o \qquad \textbf{(5.2k)}$$

$$\theta_2 = \pi/2 - \cos^{-1}\left[\frac{\mathbf{q}_2 \odot \mathbf{n}_2}{|\mathbf{q}_2| \times |\mathbf{n}_2|}\right] \qquad \text{where } -90^o \le \theta_i \le 90^o \qquad \textbf{(5.2l)}$$

(a)



Key: **tp*i*** = triangular plane *i*, where *i* = 0, 1, 2, 3.
rp*i* = reference plane *i*, where *i* = 1, 2.
Note: $\theta_i' = 180^0 - \theta_i$, for *i* = 0, 1, 3.

(b)

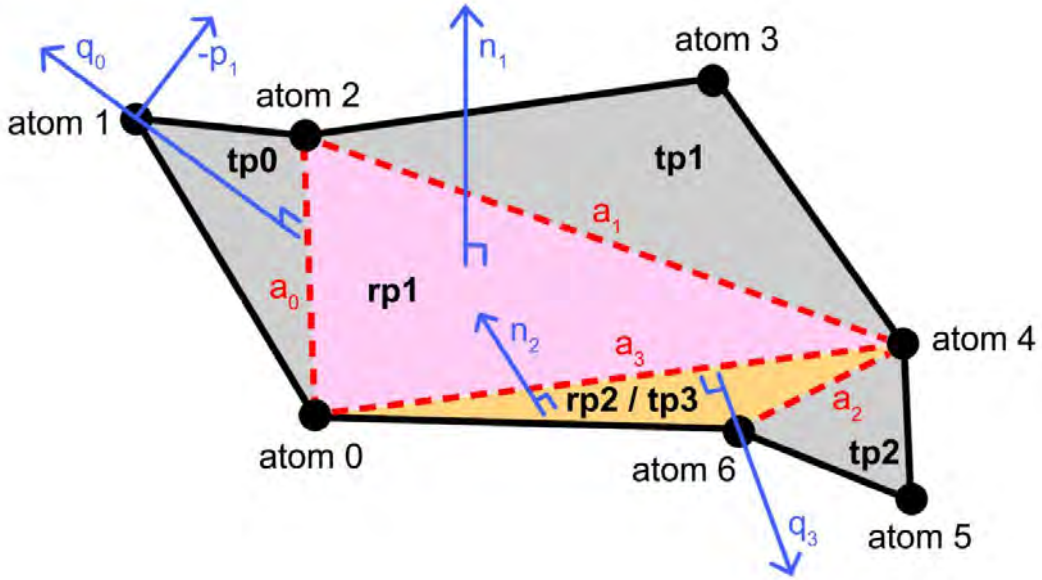


Figure 5.7: A representation of the 7-membered ring's triangular tessellation shown in the *chair* conformation. In **(a)** the pucker angles (angular puckering coordinates) are shown while in **(b)** representative vectors involved in the computation of the angular puckering coordinates (not drawn to scale) are shown.

In the case of the triangular tessellation of 8-membered rings, the derivation of the APC is outlined in Equations **5.3a – l** below. The axes of puckering, $\mathbf{a}_i$ are defined in terms of the ring atom Cartesian coordinates, $\mathbf{x}_i$, as given by Equations **5.3a – c**. In Equation **5.3a**, $\mathbf{x}_0 \equiv \mathbf{x}_8$.

$$\mathbf{x}_i = \langle x_i, y_i, z_i \rangle \qquad \text{for } 0 \le i \le 8 \qquad \textbf{(5.3a)}$$

$$\mathbf{a}_i = \mathbf{x}_{2(i+1)} - \mathbf{x}_{2i} \qquad \text{for } 0 \le i \le 3 \qquad \textbf{(5.3b)}$$

$$\mathbf{a}_4 = \mathbf{x}_0 - \mathbf{x}_4 \qquad \textbf{(5.3c)}$$

Vectors normal to **rp1** and **rp2**, namely $\mathbf{n}_1$ and $\mathbf{n}_2$ respectively, are calculated as the cross product of two axes of puckering, as shown in Equations **5.3d** – **e**.

$$\mathbf{n}_1 = \mathbf{a}_1 \otimes \mathbf{a}_0 \qquad \textbf{(5.3d)}$$

$$\mathbf{n}_2 = \mathbf{a}_3 \otimes \mathbf{a}_2 \qquad \textbf{(5.3e)}$$

Bond vectors, $\mathbf{r}_i$, are defined and used to define orientation vectors, $\mathbf{p}_i$, for **tp**s 0 to 4, given in Equations **5.3f** – **h**. The orientation vector for **tp4** is denoted by $\mathbf{p}_{\mathbf{tp4}}$, given in Equation **5.3h**.

$$\mathbf{r}_i = \mathbf{x}_{i+1} - \mathbf{x}_i \qquad \text{for } 0 \le i \le 7 \qquad \textbf{(5.3f)}$$

$$\mathbf{p}_i = \mathbf{r}_{i-1} \otimes \mathbf{r}_i \qquad \text{for } i \in \{1, 3, 5, 7\} \qquad \textbf{(5.3g)}$$

$$\mathbf{p}_{\mathbf{tp4}} = \mathbf{a}_2 \otimes \mathbf{a}_3 \qquad \textbf{(5.3h)}$$

The axes of puckering and orientation vectors are used to define the vectors $\mathbf{q}_i$ for **tp**s 0 to 3, given in Equation **5.3i**; $\mathbf{q}_{\mathbf{tp4}}$ (for **tp4** of the ring) is a special case, defined separately in Equation **5.3j**. $\mathbf{q}_{\mathbf{tp4}}$ (also referred to as $\mathbf{q}_4$) lies along the ⊥ line joining $\mathbf{a}_4$ and $\mathbf{x}_6$, depicted later in **Figure 5.8** (where $\mathbf{x}_6$ corresponds to atom 6).

$$\mathbf{q}_i = \mathbf{a}_i \otimes \mathbf{p}_{2i+1} \qquad \text{for } 0 \le i \le 3 \qquad \textbf{(5.3i)}$$

$$\mathbf{q}_{\mathbf{tp4}} = \mathbf{q}_4 = \mathbf{a}_4 \otimes \mathbf{p}_{\mathbf{tp4}} \qquad \textbf{(5.3j)}$$

The pucker angles are then calculated in Equations **5.3k** – **l**. As in the case of 6- and 7-membered rings, **tp**s positioned above and below their respective reference planes have positive and negative pucker angles, respectively.

$$\theta_i = \pi/2 - \cos^{-1}\left[\frac{\mathbf{q}_i \odot \mathbf{n}_1}{|\mathbf{q}_i| \times |\mathbf{n}_1|}\right] \qquad \text{for } i \in \{0, 1, 4\}, \text{ where } -90^o \le \theta_i \le 90^o \qquad \textbf{(5.3k)}$$

$$\theta_i = \pi/2 - \cos^{-1}\left[\frac{\mathbf{q}_i \odot \mathbf{n}_2}{|\mathbf{q}_i| \times |\mathbf{n}_2|}\right] \qquad \text{for } i \in \{2, 3\}, \text{ where } -90^o \le \theta_i \le 90^o \qquad \textbf{(5.3l)}$$

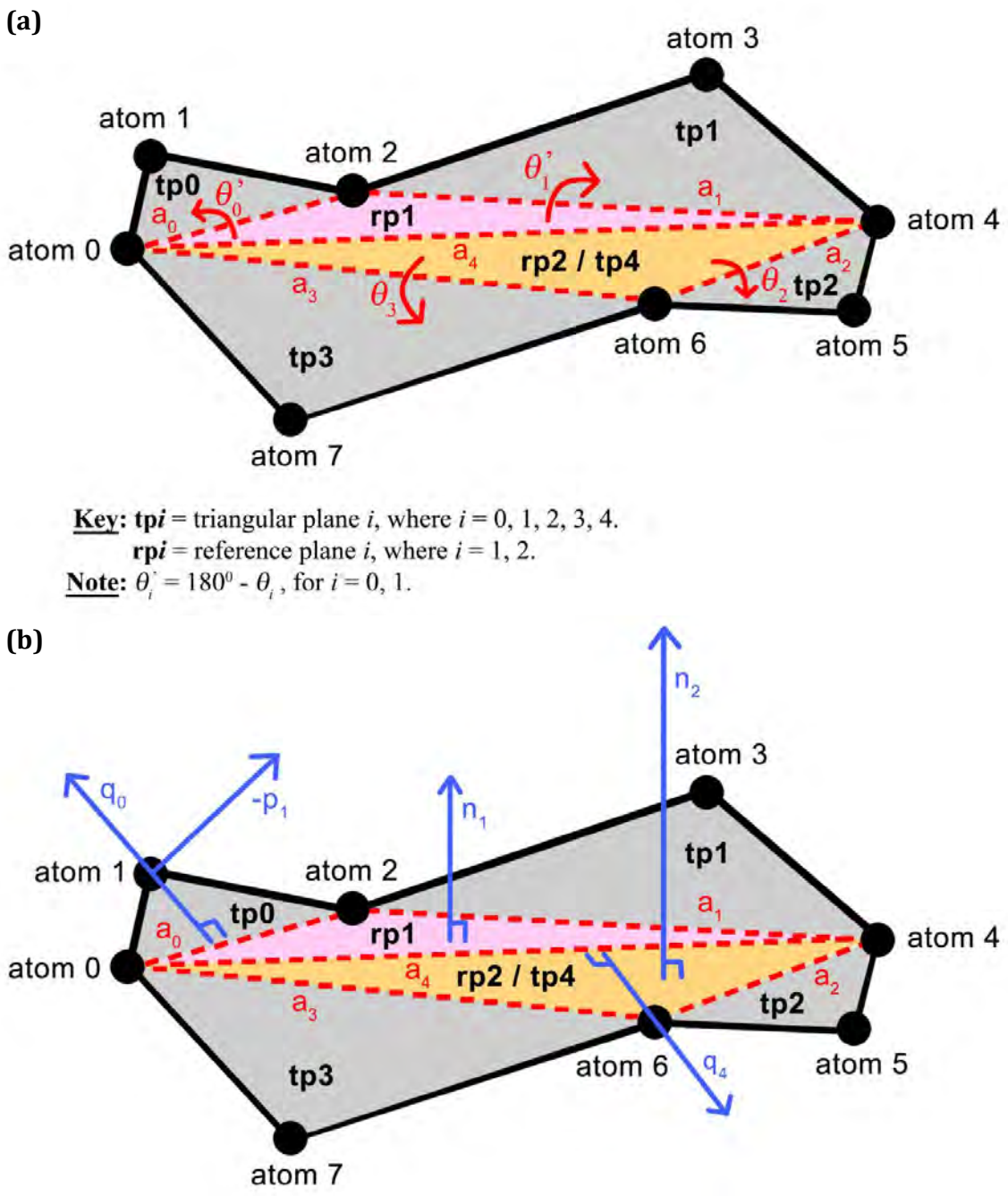


Figure 5.8: A representation of a triangular tessellation performed on an 8-membered ring, shown in the chair conformation. In **(a)** the pucker angles (angular puckering coordinates) are shown while in **(b)** representative vectors involved in the computation of the angular puckering coordinates (not drawn to scale) are shown.

5.4 Systematic Procedure to obtain APC

Previously, it had been shown for 6-membered rings that the angular puckering coordinates (APC) gained from TT can be used to define the 38 canonical states of cyclohexane [1]. Here we detail a procedure to systematically generate the APC of all the

states for 7- and 8-membered rings. Using the complete set of APC for each ring, we show how the canonical states of cycloheptane and cyclooctane may be characterised. The basic set of canonical shapes for cycloheptane (**Figure 5.1**) and cyclooctane (**Figure 5.2**) have been drawn from the Cartesian coordinates given by Wiberg [5].

The APC of all the canonical states of small and medium sized rings for hetero- and homogenous rings can be generated through the procedure described here. However here, we specifically describe the process that comprises a (i) permutation cycle, (ii) negation cycle, and (iii) removal of duplicate pucker coordinates for 6-, 7- and 8-membered rings – this is called "Procedure 1" herein. Here *one* known state is *chosen at random* from each (canonical) shape and used to generate the APC of *all* other states for that shape. For example, for the *boat* shape of 6-membered rings, one can start with the Cartesian coordinates of say, the $^{1,4}B$ *boat*, and doing Procedure 1 would generate the APC of all six *boat* states. In this setup, a set of shapes derived from ring atoms that have been numbered sequentially (i = 0, 1, 2, ..., N), starting from a randomly selected atom (i.e. labelled i = 0), and then used to identify **rp**s and **tp**s is not an absolute definition of the conformational possibilities for the ring. This is because an individual ring canonical conformation depends on the atom numbering selection. However, the complete set of APC generated for all the states of each shape is independent of the choice of atom numbering scheme. Thus the aim here is to generate APC of all the states of a ring but *not* to associate these APC with *particular* states. Procedure 1 is described as follows:

PROCEDURE 1

Consider a canonical **shape *x*** of an N-membered ring, where $N \in \{6,7,8\}$.

Permutation Cycle:

- Cartesian coordinates are read for a randomly selected state ***y*** taken from the set for **shape *x***. A start atom (i = 0) is assigned and the numbering for the ring atoms is done in a clockwise manner, from 0 to N, where atom0 is equivalent to atom$N$.

- The APC of state ***y*** are calculated from the Cartesian coordinates, using the TT approach as detailed above in (**Figures 5.3, 5.5 and 5.6**) and the equations for its APC derivation, i.e. Equations **5.1a – g**, **5.2a – l**, and **5.3a – l**.

- Following this, the atom numbers are shifted along one ring atom position in a clockwise direction around the ring. This number ***permutation*** about the ring is repeated *N* times, and in each case the APC are calculated.

  The permutations are labelled **P*i***, where $1 \leq i \leq N$. As an example, permutations **P1** and **P2** for a 6-membered ring are shown in **Figure 5.9**. Each permutation represents the APC of either a **new state** or **redundant state** of **shape *x***, where a redundant state would be one whose APC have been generated in a previous permutation.

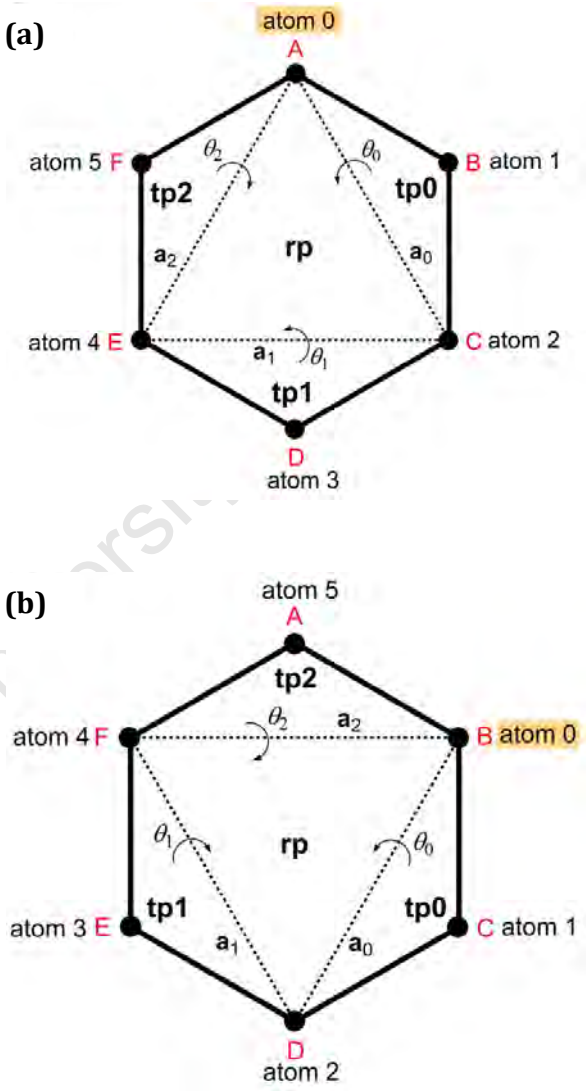


Figure 5.9: Two permutations **(a) P1** and **(b) P2** of a 6-membered ring generated from the permutation cycle of Procedure 1. The letters A to F represent the Cartesian coordinates of the ring atoms.

Negation Cycle:

The permutation cycle may not generate all the possible states of **shape *x***. The remaining states may be found using **rp**s as reflection planes. This is termed the negation cycle and is described as follows:

- For each permutation producing APC, a reflection about the **rp**s (termed a *negation*) is applied. For example, a 6-membered ring with coordinate set (θ_0, θ_1, θ_2) produces APC (-θ_0, -θ_1, -θ_2), where the latter APC are labelled **N*i***, with $1 \leq i \leq N$. The APC derived in this way are examined for their uniqueness i.e., they are either new or a repeat of coordinates generated by the preceding permutation cycle.

The discovery of APC from **rp** reflections is illustrated in **Figure 5.10** for the *boat* shape of a 6-membered ring. Having chosen a state from the *boat* shape, **Figure 5.10(c)** is any permutation of this state, and the reflection of the **tp**s about the **rp** results in a state that may be novel or a repeat (**Figure 5.10(d)**). Taking cyclohexane as the specific example of the 6-membered ring in **Figure 5.10(a)** and **(b)**, if the IUPAC numbering is such that C1 is atom 0, C2 is atom 1, and so on, then **Figure 5.10 (a,c)** is the $^{1,4}B$ state and **Figure 5.10 (b,d)** is the $B_{1,4}$ state.

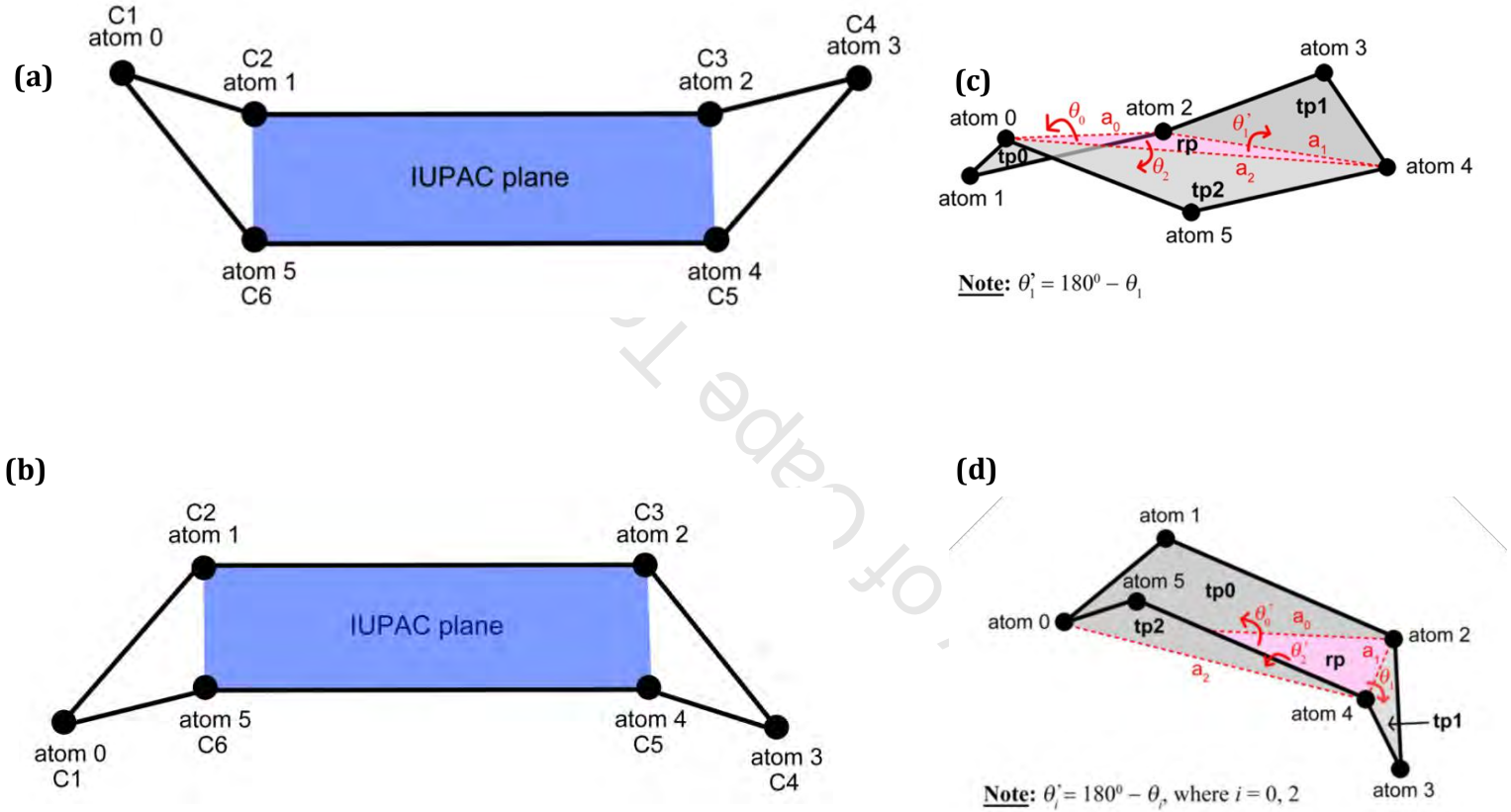


Figure 5.10: (a,c) A general permutation **P*i*** of the *boat* shape of a 6-membered ring, where (θ_0, θ_1, θ_2) = (x°, y°, z°). **(b,d)** The reflected state **N*i*** of permutation **P*i***, where (θ_0, θ_1, θ_2) = (-x°, -y°, -z°) for **N*i***. In **(a)** and **(b)**, cyclohexane is taken as the example of the 6-membered ring, where the IUPAC numbering is related to the atom numbering as C1 being atom 0, C2 being atom 1, and so on.

Removing Duplicate APC:

For the *N*-membered ring, a total of 2*N* APC are generated for each canonical shape from the permutation and negation cycles. Duplicates can be removed "on-the-fly", that is, as soon as a duplicate APC is generated it is discarded. Alternatively, once all possible APC are generated with duplicates, a simple comparison of pairs of angular puckering coordinates can be done and any duplicates removed.

5.5 Verification of Procedure 1 for 6-membered Rings

Using the endocyclic torsion angles of the 38 ideal states of pyranoses [13], Hill and Reilly [1] have generated the structures of the states of cyclohexane and calculated the APC of these states using triangular tessellation (the APC are given in **Table A1** in **Appendix A**). In **Table A1**, the APC have been assigned to *specific states* by using a particular atom numbering scheme (**Table 5.3**) to relate the IUPAC numbering of the ring (which defines the naming of the states) to its atom numbering.

Table 5.3: Atom numbering scheme used in reference [1] to generate APC for cyclohexane.

IUPAC numbering	Atom numbering
C1	atom 0 / atom 6
C2	atom 1
C3	atom 2
C4	atom 3
C5	atom 4
C6	atom 5

In this section, the validity of Procedure 1 is tested by implementing it on the cyclohexane system, and then comparing the APC generated to those listed in **Table A1**. This was carried out using a combination of previously developed and locally developed python codes, which can be accessed from the Scientific Computing Research Unit website.[a]

[a] The Scientific Computing Research Unit website can be accessed at http://www.scientificcomputing.com/

The process is outlined in the steps below:

1. *N* was set equal to 6 in Procedure 1, and **shape *x*** ∈ {*chair*, *boat*, *half-chair*, *skew-boat*, *envelope*}.
2. The APC of the 38 states of cyclohexane listed in **Table A1** were converted into Cartesian coordinates, by estimating the values of the unknown bond lengths and angles, using previously developed python code, which can be accessed from the Scientific Computing Research Unit website.[b]

3. For each **shape *x***, the Cartesian coordinates of ***all*** the states were used as starting points in Procedure 1 to generate APC – this was done to verify that *any* state of **shape *x***, chosen *at random*, can be used to generate the APC of all the other states of the shape.

    - For each state, the ring atoms were numbered according to the atom numbering scheme given earlier in **Table 5.3**.

    - The permutation and negation cycles were then carried out. The APC of the six permutations of each state were calculated using its Cartesian coordinates together with Equations **5.1a – g**. These APC were then passed through the negation cycle to generate those of the six negations. All twelve APC for each state are given in **Table A2** in **Appendix A**.

    - For each shape and its states in **Table A2**, the APC are assigned greek letters, so that duplicate APC have the same greek letter designation. Duplicate APC were removed, and the final APC for each state are given in **Table A3** in **Appendix A**.

For each shape in **Table A3**, *no matter which* state is used and passed through Procedure 1, the *same set of APC* is generated. Thus *any* state, *chosen at random*, can

[b] The Scientific Computing Research Unit website can be accessed at http://www.scientificcomputing.com/

indeed be used as the starting point for Procedure 1. Furthermore, the final sets of APC generated from a particular state (**Table A3**) constitute the APC of *all* the states of the corresponding shape (**Table A1**). For example, for the *boat* shape, the APC of **P3** of the $B_{1,4}$ state (**Table A3**) match to those of the $B_{2,5}$ state (**Table A1**). Procedure 1 has thus been verified for cyclohexane, and can also be extended to *N*-membered rings in general. A point to note is that the actual atom numbering scheme used for the ring system does ***not*** affect the APC generated, thus another scheme can also be used, for example, C2 can be assigned atom 0, C3 as atom 1, and so on.

With regards to the validation process for Procedure 1, the resolution of duplicate APC in **Table A2** is high. The APC of the permutations and negations in **Table A3** can also be matched to the conformational states in **Table A1** within high resolution. For example, the APC of the permutation $^{1}C_{4}$–**P2** in **Table A3** is (35.25°, 35.26°, 35.26°), corresponding to the APC of the $^{4}C_{1}$ state in **Table A1**, i.e. (35.26°, 35.26°, 35.26°). Another example is the APC of the permutation $^{3}H_{2}$–**P4** in **Table A3**, (-7.90°, 15.61°, 40.23°), representing the APC of the $^{6}H_{5}$ state in **Table A1**, i.e. (-9.07°, 17.83°, 42.16°).

The resolution of APC mentioned above is high but not perfect, which can be explained by considering the process of converting from APC to Cartesian coordinates (carried out in Step 2.). The system of APC represents *N*-3 degrees of freedom, which is significantly reduced from the 3*N* degrees of freedom contained within the Cartesian coordinates. At least 3*N*-6 internal coordinates (or 3*N* Cartesian coordinates) are required to specify the geometry of a ring. It is not possible to completely specify the geometry of all ring coordinates from only *N*-3 puckering coordinates. It is possible to make approximations about most-likely bond lengths and angles and in addition use the puckering coordinates to generate an estimate of a plausible geometry [1, 14].

5.6 Generating APC for 7- and 8-membered Rings

Procedure 1 was then used to generate the APC of all the states of cycloheptane and cyclooctane, by setting *N* equal to 7 and 8, respectively. This was carried out by using a modification of previously developed python code together with locally developed

python code.[c] The set of canonical shapes used for these rings are those depicted in **Figures 5.1** and **5.2**, and tabulated in **Tables 5.1** and **5.2**. The optimised Cartesian coordinates for these canonical shapes [5] are those of one of the states of that shape – these coordinates were used in Procedure 1 to generate the APC, which are given in **Tables B1** and **C1** in **Appendices B** and **C**, respectively. From these tables, there are a total of 70 and 114 different states for cycloheptane and cyclooctane, respectively. The APC can also be used to characterise 7- and 8-membered rings in general.

5.7 Conclusion

APC gained from the TT method are useful for *quantitatively* monitoring the conformational changes that ring systems undergo in, for example, computer simulations, and correlating these changes with macrocyclic physical and chemical properties. For example, the *range* of ring puckering can be measured by monitoring the change in APC over time. Further, the *rate of decay* of the puckering motion may be computed from the *time correlation functions* (TCFs) of the APC. Puckering TCFs can give information on the flexibility of ring systems, which can in turn be related to macroscopic properties of the system, such as experimentally-observed solubility trends. In addition to simple monocyclic rings, APC can be computed for macrocyclic rings, such as cyclodextrins and porphyrins, by defining the macrocylic ring in terms of monomer repeat units making up the ring.

It is also desirable to relate the quantitative measures to qualitative descriptions of pucker that are more meaningful to scientists. For example, **Tables A1, B1** and **C1** relate the APC of cyclohexane, cycloheptane and cyclooctane to their *ideal* canonical shapes, and can be used in theory to assign the shapes of 6-, 7- and 8-membered rings based on the APC. This is relatively simple to do for static structures, where the APC can be used to assign the structure to the closest ideal conformation/s. However, assigning whole dynamics trajectories can be difficult, where the ring can take on several conformations intermediate to the ideal forms. In addition, different *extents* of puckering and strained conformations are also possible. In these cases, the APC of the ring do not exactly match or are not particularly close to an ideal form, thus assigning

[c] The Scientific Computing Research Unit website can be accessed at http://www.scientificcomputing.com/

conformations here becomes very subjective, i.e. at what point is the ring still a particular shape such as a *chair* or *boat*, and how does one decide which shape to assign it to.

One approach is to calculate the Euclidean distance (Equation **5.4**) between the APC of the ring and the APC of all the *ideal* canonical states, and then assign the ring to the canonical form for which the distance is the shortest. However, assigning to a *single* conformation may not always be accurate for reasons discussed in the previous paragraph. Equation **5.4** is given below, where *N* is the number of ring atoms, and x_i and x_i' are the pucker angles of the ring under study and the canonical form, respectively.

$$Euclidean\ distance = \sqrt{\sum_{i=0}^{N-3} (x_i - x_i')^2} \qquad \textbf{(5.4)}$$

Another approach is to express intermediate conformations as a linear combination of the ideal forms [9, 15]. Here, Equation **5.4** is used to calculate the deviations of the ring's pucker coordinates from those of ideal forms, in order to get the relative contributions of these forms to the overall ring conformation. Equation **5.4** was used with pucker coordinates based on the CP model to analyse the conformations of some 8-membered rings [9]. This method however also has its limitations and should only be used as a guide to compare the relative contributions of the canonical forms [9], rather than as a definite assignment of the conformation.

References

1. Hill, A. D.; Reilly, P. J. *Journal of Chemical Information and Modeling* **2007**, *47*, 1031.
2. International Union of Pure and Applied Chemistry and International Union of Biochemistry and Molecular Biology Joint Commission on Biochemical Nomenclature (JCBN) *Pure and Applied Chemistry* **1996**, *68*, 1919.
3. Bocian, D. F.; Strauss, H. L. *Journal of the American Chemical Society* **1977**, *99*, 2866.
4. Ivanov, P. M.; Ōsawa, E. *Journal of Computational Chemistry* **1984**, *5*, 307.
5. Wiberg, K. B. *Journal of Organic Chemistry* **2003**, *68*, 9322.
6. Hendrickson, J. B. *Journal of the American Chemical Society* **1967**, *89*, 7036.
7. Rocha, W. R.; Pliego, J. R.; Resende, S. M.; Dos Santos, H. F.; De Oliveira, M. A.; De Almeida, W. B. *Journal of Computational Chemistry* **1998**, *19*, 524.
8. DeMatteo, M. P.; Snyder, N. L.; Morton, M.; Baldisseri, D. M.; Hadad, C. M.; Peczuh, M. W. *Journal of Organic Chemistry* **2005**, *70*, 24.
9. Evans, D. G.; Boeyens, J. C. A. *Acta Crystallographica Section B-Structural Science* **1988**, *44*, 663.
10. Strauss, H. L.; Pickett, H. M. *J Am. Chem. Soc.* **1970**, *92*, 7281.
11. Joshi, N. V.; Rao, V. S. R. *Biopolymers* **1979**, *18*, 2993.
12. Bocian, D. F.; Pickett, H. M.; Rounds, T. C.; Strauss, H. L. *J. Am. Chem. Soc.* **1975**, *97*, 687.
13. Bérces, A.; Whitfield, D. M.; Nukada, T. *Tetrahedron* **2001**, *57*, 477.
14. Zou, W.; Izotov, D.; Cremer, D. *The Journal of Physical Chemistry A* **2011**, *115*, 8731.
15. Cremer, D.; Dick, B.; Christeu, D. *Journal of Molecular Structure: THEOCHEM* **1984**, *110*, 277.

Chapter 6

Coarse grain analysis of α-, β- and γ-cyclodextrin macrocyclic ring dynamics in water

The conformational behaviour of cyclodextrins (CDs) and their derivatives have been the focus of both experimental (X-ray, NMR) [1, 2] and computational studies [3-7]. On the basis of early X-ray studies, it was initially thought that CDs had a rigid structure both in the solid state and in solution. However, NMR studies as well other experimental and computational results revealed that CDs are in fact rather flexible molecules, both in solution and the solid state, and that CD complexes in solution are dynamic systems that exist in equilibrium with the free uncomplexed CD molecules [1]. For example, the crystal structure of a hydrate of a β-CD derivative showed conformational flexibility in the molecule [2], and solution studies on small-ring CDs in water revealed that their flexibility arises from different internal motions in the molecules themselves [3, 6], as well as from interactions with the solvent water molecules [6]. Investigations into large-ring CDs have also been done. For example, the conformations of CDs with 24 to 30 glucose units have been studied through MD simulations together with a statistical method called Principal Component Analysis (PCA) [8, 9], where it was found that the CDs often formed loops and helical fragments as favourable deformations, reinforcing the idea that they are flexible molecules.

6.1 Conformational Behaviour of Cyclodextrins

The focus of this chapter is to investigate the relative conformational flexibilities of α-, β- and γ-CD in water. I will then seek to understand if there is a correlation between the dynamic ring conformational motion and the solubility trend for these CDS (γ-CD > α-CD ≫ β-CD). Section 6.2 presents an overview of the conformational behaviour of CDs, as investigated through different metrics. Section 6.3 gives the computational details of the simulations in this work, and Section 6.4 attempts to explain the solubility trend of α-, β- and γ-CD through a coarse grain analysis of the macrocyclic rings. This analysis involves (i) measuring the macrocyclic puckering of the CD rings from MD trajectories in order to investigate their relative flexibilities, and (ii) measuring the CD "breathing"

motion (planar deformation motion of the macrocyclic rings) from the trajectories, to supplement the pucker analysis.

6.2 Conformational and Configurational Analysis

Different metrics have been used in literature to study the conformational behaviour of small and medium sized CD molecules (with 6 – 9 glucose units) and their derivatives [3, 5-7, 10-12]. On the macrocyclic scale, distortions of the macrocyclic CD ring as a whole are investigated by studying motions such as rotations about the glycosidic bonds. On the monomeric scale, motions such as the horizontal tilting, vertical librations and twisting of the individual glucose monomers have been studied. The presence and extent of intramolecular hydrogen bonding in the CDs also contributes to their overall motions and interactions with water. Several hydration studies have also been carried out to investigate how the CD motions affect the structure and behaviour of the surrounding water molecules. These factors, as well as their relations to the flexibilities and aqueous solubilities of CDs, are discussed in more detail below.

6.2.1 Macrocyclic and Monomeric Motions

An important macrocyclic motion in CDs is the rotation about the glycosidic bonds between adjacent glucose units in the molecules. Computer simulations (both in vacuo [4] and in aqueous solution [3, 6]), combinations of computational and NMR techniques [10], as well as crystal structures [11] have all been used to measure the glycosidic bond angles and study the glycosidic rotations. The rotations are quantified by measuring changes in the dihedral angles of rotation ϕ and ψ about the glycosidic bonds (see **Figure 6.1**). The rotations can then be analysed by plotting probability distributions [3, 10] and TCFs [3] of the ϕ and ψ angles. In **Figure 6.1,** the ϕ and ψ angles are defined as ϕ (H1' – C1' – O4 – C4) and ψ (C1' – O4 – C4 – H4) [3], where " ' " indicates atoms of the adjacent glucose monomer, but there are also other slightly varied definitions, where hydrogens H1' and H4 of the ϕ and ψ angles are replaced by O5' and C5 [11], and O5' and C3 [6], respectively. Although the ring structure of the CD limits the extent of rotation about the glycosidic bonds, as compared to maltose (i.e. glucose dimer) [3], the rotations are still an important factor for CD conformational flexibility.

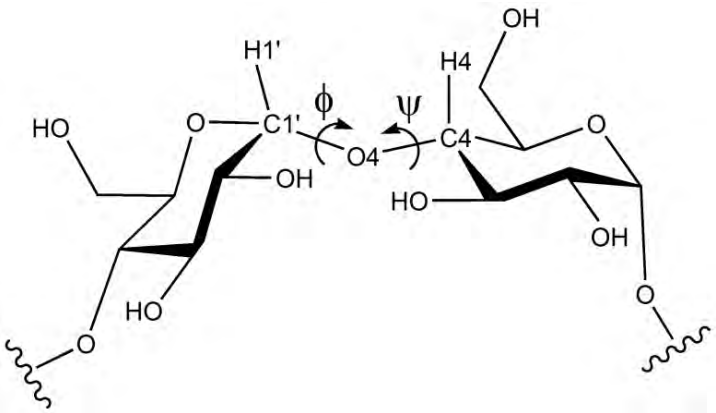


Figure 6.1: Schematic representation of the glycosidic bond rotations ϕ and ψ in CDs (as defined in reference [3]), shown here for a portion of the molecule. ϕ and ψ are dihedral angles defined by the atoms (H1' – C1' – O4 – C4) and (C1' – O4 – C4 – H4), respectively.

Different studies have also been carried out on the monomeric scale, investigating motions such as the degree of ring twisting of the glucose units in CD crystal structures (measured about a pseudo torsion angle passing diagonally through the glucose units) [11], horizontal / angular tilting of the glucose units towards or away from the macrocyclic plane (which measures the relative orientations of the glucose units to the CD ring) [3, 13], and vertical displacements / librations of the glucose units relative to the macrocyclic plane [3] – the two latter motions are illustrated in **Figure 6.2**.

(a) **(b)**

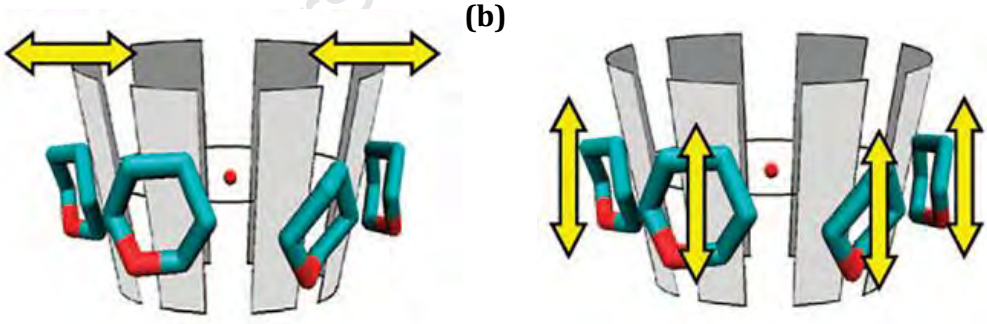


Figure 6.2: Schematic representation of the **(a)** horizontal tilting, and **(b)** vertical librations of glucose units in cyclodextrins, taken from reference [3].

The *relative* conformational flexibilities of CDs have also been investigated using the above-mentioned techniques. In 2008, Naidoo *et al*. [3] compared the relative flexibilities of α-, β-, and γ-CD in water by running 30 ns MD simulations on these systems and analysing various motions of the CDs from the trajectories. Glycosidic bond rotations were calculated through the ϕ and ψ angles (with ψ being the main

contributor to the relative CD flexibilities), the horizontal tilting motion (**Figure 6.2(a)**) of the glucose units was measured through a suitably defined tilt angle, and the librations (**Figure 6.2(b)**) were measured as vertical displacements of the centres of mass of the glucose units from the macrocyclic plane. Average TCFs of the tilt angle and ψ angle, and average RMSD TCFs of the vertical displacements all gave correlation times, τ, in the order $\tau_\gamma > \tau_\alpha > \tau_\beta$ for the respective motions, i.e. β-CD relaxes the fastest to equilibrium, followed by α-CD and then γ-CD [3]. The order of the τ values reflects the relative conformational flexibilities of the CDs in the order γ-CD > α-CD > β-CD – this mirrors the solubility trend, which shows a correlation between conformational flexibility and solubility of these three CDs (as the flexibility of these CDs increases, so does their aqueous solubilities). The *amplitude* of the vertical libration motion (measured from the RMSD TCFs) was also found to be smallest for β-CD, indicating the macrocyclic motions of β-CD to be the most restricted [3]. Probability distributions of the ϕ and ψ angles were also generated, which revealed β-CD to be a little more limited in its range of motion than the other two CDs [3], supporting the idea that β-CD is the least flexible.

Tilt angle distributions were also calculated in reference [3], the results of which agree well with later experimental results of Bernatowicz *et al*. [13], who also determined the tilt angles of α-, β- and γ-CD in their study of the Carbon-13 NMR relaxation processes of CDs (with 6 to 12 glucose units) – this shows agreement between computational and experimental methods.

6.2.2 Intramolecular Hydrogen Bonding

Another metric used to study CDs is the presence of intramolecular hydrogen-bonding in the CD molecules [3, 4, 6, 7, 14]. The main type of intramolecular hydrogen-bonding occurs between the secondary hydroxyl groups of adjacent glucose units (i.e. between the OH groups on C3 and C2' across a glycosidic bond), where the distances between these hydroxyls are sufficiently small to allow for hydrogen bond formation. This type of hydrogen bonding is depicted in **Figure 6.3**, and has been well-studied and well-established, both in solution [3, 6, 12] and in solid state crystal structures [11], as well as from NMR spectroscopy [15].

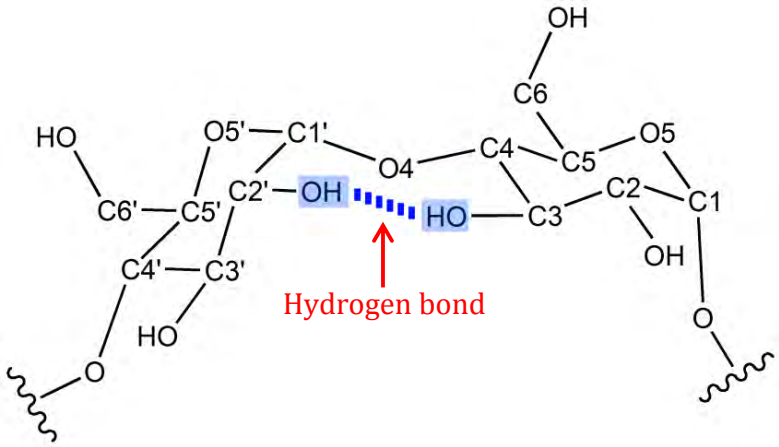


Figure 6.3: Schematic representation of the main type of intramolecular hydrogen bonding in CDs, adapted from reference [3] and shown here for a portion of the molecule. The carbon atoms are numbered C_i and C_i', where $1 \leq i \leq 6$ and " ' " indicates the atoms of an adjacent glucose unit. The hydrogen bond (dashed blue line) occurs between the secondary hydroxyl groups (highlighted in blue) of adjacent glucose units.

Different studies have investigated the effect of intramolecular hydrogen bonding on CD conformational flexibility. In an MD study of methylated β-CD derivatives, it was found that as the degree of methylation increased, the intramolecular hydrogen bonds decreased and the flexibility of the CD consequently increased [14] – thus for these CDs, the trend is that as the number of intramolecular hydrogen bonds decreases, the CD flexibility increases. Probability plots of the intramolecular hydrogen bonds in α-, β- and γ-CD in water have also been generated [3], revealing that β-CD (the least flexible of the three CDs) shows the most hydrogen bonding. The relation to solubility has also been studied. For example, Shikata *et al*. [12] used dielectric relaxation measurements to study how chemical modifications of the hydroxyl (OH) groups of α-, β-, and γ-CD change the hydration states and solubilities of the native CDs. In permethylated α-CD (pmα-CD) for example, all the secondary and primary OH groups on α-CD are changed to methoxy (OCH_3) groups, which effectively removes the intramolecular hydrogen bonding (as methoxy groups cannot form hydrogen bonds to each other) and increases the hydration number of pmα-CD relative to α-CD, thereby increasing the solubility of pmα-CD – this is depicted in **Figure 6.4** [12]. Thus for the CDs studied in reference [12], as the intramolecular hydrogen bonding is lessened, the solubility of the CD increases. In **Figure 6.4**, the values of the hydration numbers of the primary and secondary hydroxyl groups of α-CD are *one*, indicated by the *single* water molecules hydrogen-bonded to the hydroxyl groups. Similarly, the hydration numbers of the methoxy groups

in pmα-CD are *four*, indicated by the *four* water molecules hydrogen-bonded to the methoxy groups [12]. The reaction arrow in **Figure 6.4** indicates the permethylation process and the "addition" of water molecules, which are indicated in the figure as per reference [12].

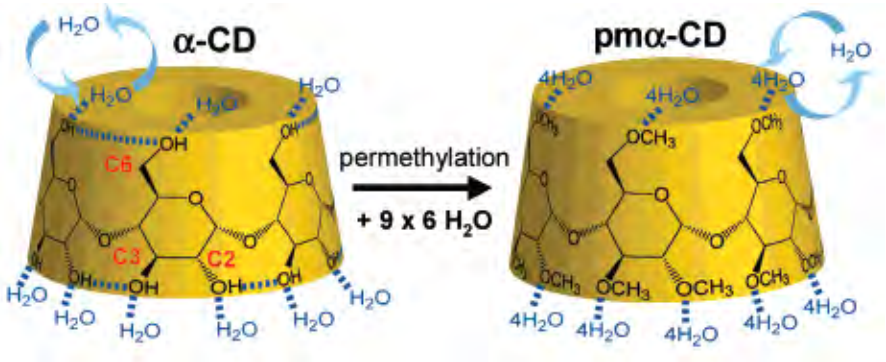


Figure 6.4: Permethylation of α-CD to pmα-CD, taken from reference [12] (the relaxation time τ_2 has been omitted from the original figure). The OCH_3 groups in pmα-CD interact with more water molecules than the OH groups in α-CD, due to the absence of intramolecular hydrogen bonding in pmα-CD (dotted lines are hydrogen bonds).

6.2.3 Hydration Studies

In aqueous solutions, both the primary and secondary hydroxyls on the CDs can form intermolecular hydrogen bonds with the surrounding water molecules [16, 17]. This in turn influences the dynamical behaviour of the water molecules. Several computational studies have investigated the dynamics and structure of the solvation shells around the CDs [16, 18, 19], as well as the distribution of water molecules around the hydroxyl oxygen atoms of the glucose units [14]. The residence time and distribution of water molecules in the CD cavities, as well as their diffusion into the cavities, have also been explored [6, 7, 20, 21].

A recent MD study on α-, β-, and γ-CD in water [5] revealed the structuring of water molecules around the CDs and in the CD cavity by generating three-dimensional (3D) spatial distribution functions (SDFs) from the MD trajectories, as depicted in **Figure 6.5**. The SDFs show the regions around the CDs where the probability of finding water molecules is 50% higher than in bulk solvent. From the SDFs, the least soluble *β*-CD imposes the greatest structuring or ordering of the local water molecules (both inside the CD cavity and around the molecule), while γ-CD, being the most soluble of the three

CDs, structures the surrounding water the least. α-CD has intermediate solubility and its SDF is also intermediate between β- and γ-CD [5]. The results of the SDFs are also consistent with previous studies on the enthalpy and entropy of solution of these CDs [19, 22], which noted that the decreased solubility of β-CD was linked to its unfavourable entropy of solution. Thus one can see a correlation between the solubility of these three CD systems and their structuring of the local water molecules – the more order is imposed on the water molecules (i.e. the less the entropy of solvation), the less soluble the CD. The structuring of the water can also be linked to the flexibility of the macrocyclic ring, i.e. the more restricted motions of the β-CD ring means that it is not as easily solvated or incorporated into the water structure as α- and γ-CD [5].

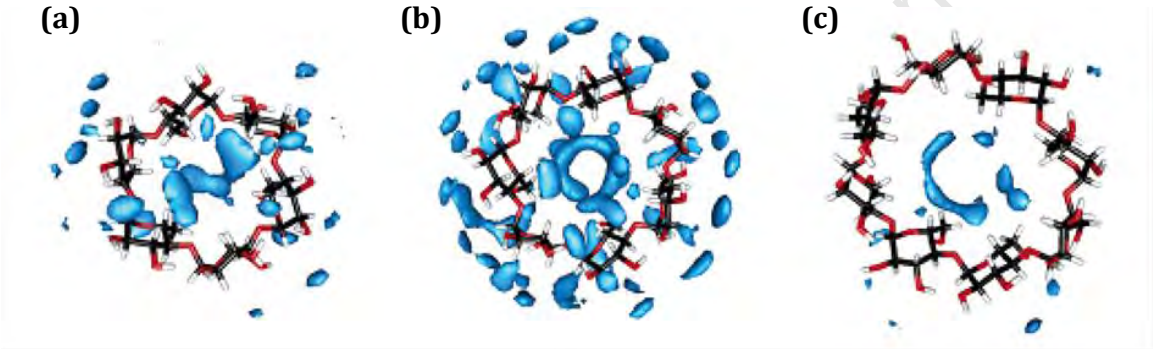


Figure 6.5: Spatial distribution functions (SDFs) of **(a)** α-, **(b)** β- and **(c)** γ-CD in water, taken from reference [5].

The link between solubility and the hydration states of chemically modified CDs has also been studied. For example, as mentioned previously, Shikata *et al*. [12] investigated how chemical modifications of α-, β- and γ-CD impacted on the hydration numbers, hydrogen-bonding sites, and solubilities of the modified CDs.

6.3 MD Simulation Protocol

In a previous study by Naidoo *et al.* [3], 30 ns MD simulations have been run on α-, β- and γ-CD in water using the CSFF force field [23] and the SPC/E water model [24], where both the force field and water model were found to be useful in studying the conformational behaviour of the CDs. In particular, the study compared the coupling constants calculated for the CDs from the simulations [3] to those obtained from NMR data [10], from which it was concluded that the simulations were long enough and the force field accurate enough to study the conformational space accessed by α-, β- and γ-

CD [3]. The present work expands on this study [3], where the same simulations are extended for a longer time period for the puckering and breathing analyses here, thus naturally, the same force field and water model are used here. Other carbohydrate force fields have been developed in recent years that incorporate parameters for carbohydrates [25-28], but these are not considered here for the reasons: (i) as mentioned above, this work expands on the work of reference [3], thus the same force field is used here, and (ii) this work involves a *coarse grained* analysis of full-atomistic simulations, where the *macrocyclic* CD ring puckering is analysed rather than the *local* puckering in the CDs, i.e. the focus of this work is not on the *atomistic* detail of the simulations, in which case further consideration to the choice of force field may have been required.

The 30 ns simulations from reference [3] are extended herein to 50 ns using the same simulation setup and conditions. Subsequent analysis is based on the 50 ns trajectories. The simulation conditions are summarised below.

The MD simulations were carried out in CHARMM [29] and the CDs were modelled using the CSFF force field [23]. A cubic simulation box with dimensions of 49.6105 Å on each side was used with PBC to simulate the systems, with each box containing one CD molecule and 4040 water molecules. The SPC/E water model [24] was used to characterise the water molecules. The chemical bonds involving hydrogen atoms were fixed using the algorithm SHAKE [30]. The simulations were run using the NPT (isothermal-isobaric) ensemble, where the pressure and temperature were set to constant values of P = 1 bar and T = 300 K. The Langevin Piston method [31] was used in CHARMM to keep the pressure and temperature constant, where the piston mass and piston collision frequency were set to 500 amu and 5 ps^{-1}, respectively, ensuring little effect on the molecular motions of heterogeneous systems like the cyclodextrin-water solutions being studied here [5]. The CD systems were equilibrated for 500 ps, after which 30 ns dynamic trajectories were generated, using an integration time step of 1 fs [3]. The dynamic trajectories were then run for a further 20 ns in this work to get a total length of 50 ns. The Leap-Frog Verlet algorithm [32] was utilised for the equilibration and dynamics runs, and the configurational coordinates of the system were saved in 0.05 ps intervals (i.e. after every 50 time steps). A switching function was used for the

potential energy, where the cuton, cutoff, and neighbour cutoff distances were set to 10 Å, 12 Å, and 14 Å, respectively.

6.4 Coarse Grain Analysis of Macrocyclic Cyclodextrin Rings

In this chapter, the macrocyclic ring dynamics of α-, β- and γ-CD in water are used to explain the anomalous solubility of β-CD. Two macrocyclic motions are defined and analysed from MD simulations, namely (i) the ring puckering of the CD molecules as a whole, and (ii) the "breathing" motion of the CDs (planar deformation motion of the macrocyclic rings). These two motions are investigated through a *coarse grain* analysis of the full atomistic MD simulations, as explained below.

In the CD ring, each glucose monomer and its adjacent monomer (going in a clockwise direction) are grouped together in a coarse grained manner, where each bead is made of a single *dimer unit*. There are *N* dimer units in the CD ring, where *N* is the number of glucose units in the CD. The main contribution to the overall CD ring puckering is due to torsional motion (or bond rotations) about the glycosidic bonds between the glucose units in the ring – this idea is supported in different references considering the glycosidic bond rotations [1, 3, 6]. The intermolecular interactions (hydrogen bonds) between the glucose unit substituents and surrounding water molecules also play a role in the puckering, but this is not addressed here. Thus each dimer unit is comprised only of the *ring* atoms of two glucose units and the glycosidic oxygen joining the units, i.e. all substituents on the glucose units (hydrogen, hydroxyl and primary alcohol groups) are ignored. The dimer unit is shown schematically in **Figure 6.6(a)**, and the collection of dimer units within the CD ring is shown in **Figure 6.6(b)** for α-CD. The two macrocyclic motions of ring puckering and "breathing" are then measured from *monomeric* rings defined by the centres of mass of the coarse grain beads (dimer units), as illustrated in **Figures 6.6(c)** and **(d).** The analysis of the macrocyclic puckering and "breathing" motions are considered in Sections 6.4.1 and 6.4.2, respectively.

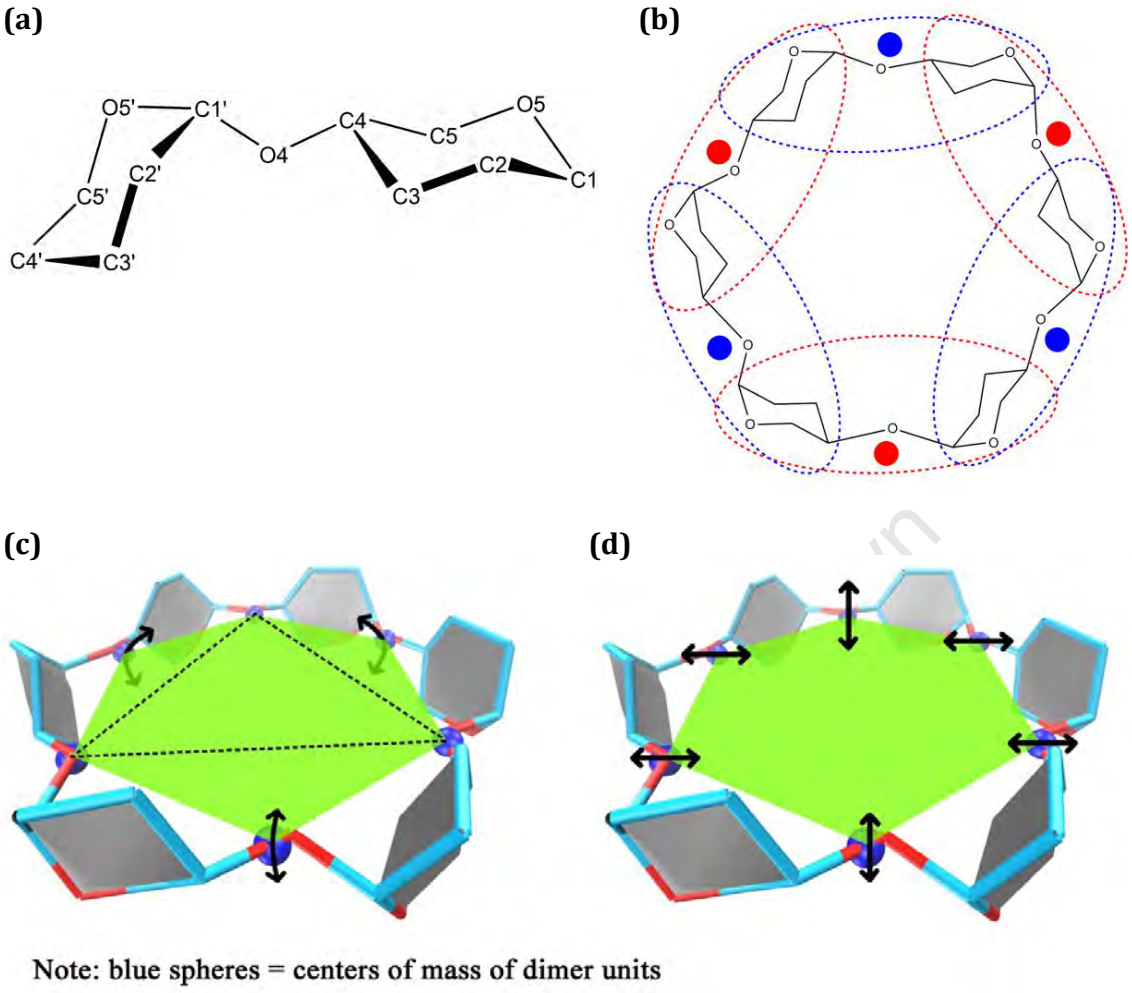


Figure 6.6: (a) Schematic representation of the coarse grain bead (dimer unit) used to analyse the motions of α-, β- and γ-CD, in this work. The carbon atoms are numbered C_i and C_i', where $1 \leq i \leq 6$ and " ' " indicates the atoms of an adjacent glucose unit. **(b)** The positions (Cartesian coordinates) of the centres of mass of the coarse grain beads (dimer units, i.e. maltose units) are then calculated and used to define the vertices of a *monomeric N*-membered ring. In calculating the centres of mass of the coarse grain bead, *only* the glucose ring atoms and glycosidic oxygens are used (the use of two colours has no physical meaning and is included to improve clarity). **(c)** and **(d)** represent the macrocyclic motions in cyclodextrins, shown here for α-CD: **(c)** shows the ring puckering motion of the cyclodextrin molecule as a whole (the dotted black lines represent imaginary axes), and **(d)** shows the planar deformation motion, or horizontal contraction and expansion ("breathing") of the cyclodextrin ring.

6.4.1 Puckering Motion

The *macrocyclic* full atomistic rings of α-, β- and γ-CD are first reduced to coarse grained *monomeric* ring systems composed of the centres of mass of the dimer units. The triangular tessellation (TT) of these monomeric rings is then used to analyse the

macrocyclic puckering of the CDs, where the TT method for 6-, 7- and 8-membered rings (as explained in Chapter 5) is used for the monomeric rings corresponding to α-, β- and γ-CD, respectively. This process is illustrated for these CDs in **Figure 6.7**. In the case of α-CD, a 6-membered monomeric ring is defined from the six dimer units. β-CD and γ-CD are simplified in a similar manner to 7- and 8-membered monomeric rings, respectively. The associated angular puckering coordinates (APC) for the monomeric rings are then used as a quantitative measure of the macrocyclic ring puckering.

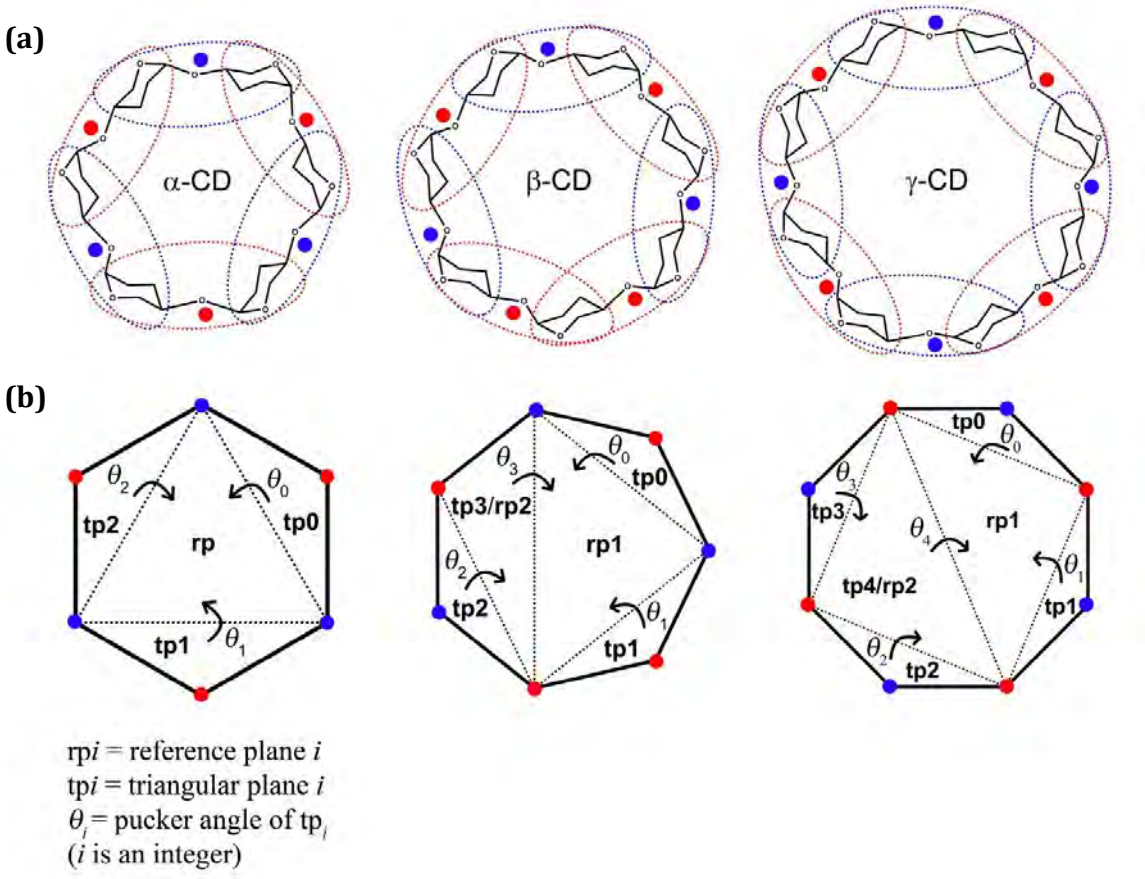


Figure 6.7: Reducing the macrocyclic full atomistic α-, β- and γ-CD rings to coarse grained monomeric ring systems. **(a)** shows only the atoms used to compute the centres of mass of the coarse grain beads, where the dimer units are circled by dotted lines. The red and blue dots represent the centres of mass of the coarse grain beads (the use of two colours has no physical meaning and is included to improve clarity). **(b)** shows the triangular tessellation of the monomeric rings representing α-, β- and γ-CD, which are used in the analysis of the macrocyclic ring conformational dynamics.

In reality, all the various motions in the CD ring (motions of glucose units, glycosidic bond rotations, macrocyclic puckering etc.) are interdependent and it can be difficult to isolate a particular motion to analyse. Previously, both the monomeric motions (e.g.

horizontal tilting and vertical librations) and glycosidic bond rotations have been isolated and studied through probability distribution plots and TCFs [3], and have a definite and measureable contribution to the overall CD motion. In this work, the aim is to try and isolate the puckering motion of the macrocyclic ring as a whole. By grouping glucose dimer units (i.e. maltose) into coarse grain beads, the monomeric motions of individual glucose units and the rotations about the glycosidic bonds are incorporated into the beads, hence the measurement of just the macrocyclic conformational dynamics is increased.

Alternative coarse graining definitions can be used to measure the dynamics of macrocyclic puckering. An obvious example would be to use the centres of mass of the glucose units. However, this choice of coarse grain bead would include the distortions of the individual glucose units in addition to the macrocyclic ring puckering. Another definition could be the use of the glycosidic oxygen atoms themselves, but this would include the motional contribution of the glycosidic bond rotations. A comparison of these three choices for coarse grain bead definition reveals that the dimer units, as shown in **Figure 6.7(a)**, is the best approach that will produce a relatively isolated measure of macrocyclic puckering.

Pucker time series

Having defined APC to measure the macrocyclic puckering of the CDs, as illustrated in **Figure 6.7**, the time series of the APC (called the *pucker time series* herein) were generated from the CD trajectories. The APC for the three CDs are summarised in **Table 6.1** below.

Table 6.1: APC of α-, β- and γ-CD, based on the definition in **Figure 6.7**.

Cyclodextrin	APC	Number of APC
A	θ_0, θ_1, θ_2	3
B	θ_0, θ_1, θ_2, θ_3	4
Γ	θ_0, θ_1, θ_2, θ_3, θ_4	5

To generate the pucker time series, dummy atoms were first placed on the calculated centres of mass of the dimer units for every frame (i.e. every 0.05 ps) in the 50 ns CD

trajectories. Using the Cartesian coordinates of the dummy atoms, the APC for every frame were calculated according to Equations **5.1a – g**, **5.2a – l**, and **5.3a – l** for α-CD, β- and γ-CD, respectively. There are three, four and five pucker time series for α-, β- and γ-CD, respectively, one for each of the APC (the time series are shown in **Figures D1**, **D2** and **D3** in **Appendix D**).

For each CD over the 50 ns trajectories, the probabilities of each pucker angle were calculated *separately* using intervals of 2^0 in the range (-100°, 100°). The probabilities of the individual APC, $\boldsymbol{P}_{\theta_i}$, were then averaged for each interval to give the average probability, $\boldsymbol{P}_{ave}$, for the interval – this is given by Equation **6.1** below:

$$\boldsymbol{P}_{ave} = \frac{\left(\sum_{i=0}^{N-4} \boldsymbol{P}_{\theta_i}\right)}{N-3} \qquad (\mathbf{6.1})$$

In Equations **6.1** (and later in Equations **6.2** and **6.3**), N is the number of glucose units in the CD. A probability distribution of the average probabilities of the pucker angles is given in **Figure 6.8** for the CD trajectories, where the average probabilities are plotted against the midpoint of each interval – for example, the average probability of the pucker angles lying in the interval (20°, 18°) is plotted against θ = 19°.

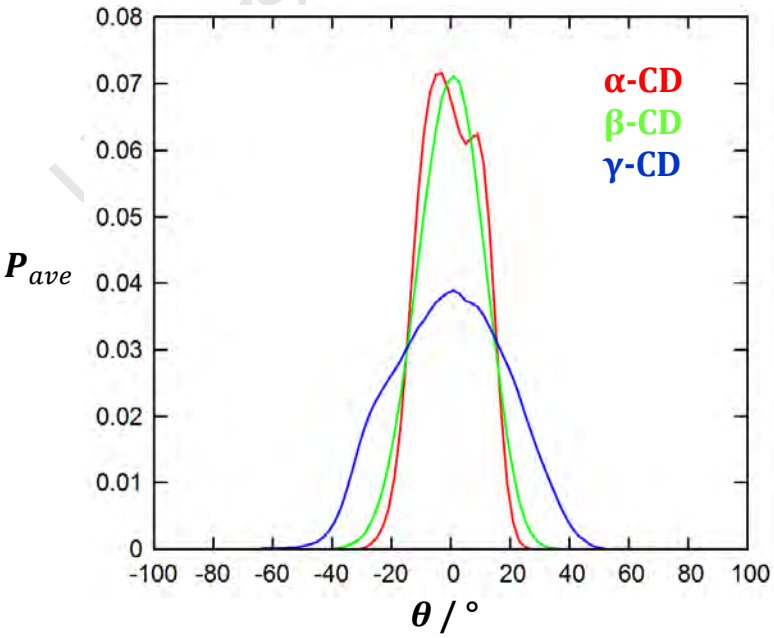


Figure 6.8: Probability distribution of the average probabilities of the pucker angles of α-, β- and γ-CD over their 50 ns MD trajectories.

Average probabilities of the angular puckering coordinates were calculated in a similar manner as calculated for the average probabilities of the tilt angle in reference [3]. This was done in addition to the pucker time series to establish the *range of motion* of the CD. The puckering behaviour across the CD molecule is expected to be similar for the different pucker angles (due to the long simulation time, as well as the fact that the CD rings are homogeneous) – this allows for the probabilities of the individual pucker angles to be averaged. A point to note about **Figure 6.8** is that the highest values of $\boldsymbol{P}_{ave}$ for the three CDs corresponds approximately to the region between (-10°, 10°) on the x-axis, indicating that the *individual* TT triangular planes have, on average, a higher probability for low-amplitude rotations relative to their respective reference planes. The probability distributions follow Gaussian-like distributions, where the means and associated standard deviations (SDs) are 0 ±10°, 0 ±11°, and 0 ±19° for α-, β- and γ-CD, respectively. The relative ranges of the distributions are 66°, 90°, and 136° for α-, β- and γ-CD respectively. Both negative and positive pucker angles are taken on by the CDs. Both the SDs and ranges follow the order α-CD < β-CD < γ-CD, where the SD and range of γ-CD are the highest of the three CDs, indicating that γ-CD has the widest range of puckering motion – this can be expected as γ-CD has the largest ring size and thus the TT triangular planes can take on a wider range of angles. The *range* of the puckering motion does *not* however reflect the relative conformational flexibilities of the rings – this will be addressed in the TCFs that follow.

Pucker TCFs

TCFs of the pucker time series were then generated for each CD, correlating over the full 50 ns trajectories. For each CD, separate TCFs were calculated for each APC using the definition in Equation **3.8** of Chapter 3. This is given in Equation **6.2** below:

$$C_{\theta_i\theta_i}(t) = \frac{\langle \delta\theta_i(t)\delta\theta_i(0)\rangle}{\langle \delta\theta_i(0)\delta\theta_i(0)\rangle} \qquad (\mathbf{6.2})$$

where $0 \leq i \leq N-4$. The TCF values for the APC at each value of t were then averaged to give the average TCF value, $C_\theta(t)$, as shown in Equation **6.3**:

$$C_\theta(t) = \frac{\left(\sum_{i=0}^{N-4} C_{\theta_i\theta_i}(t)\right)}{N-3} \qquad (\mathbf{6.3})$$

where $0 \leq i \leq N-4$. The average TCFs of the CDs are plotted in **Figure 6.9** – the plots are correlated over 50 ns but are truncated at t = 2 ns in **Figure 6.9(a)** since their decay reaches zero within this time. The pucker TCFs are shown up to t = 20 ns (about 1/3 of the trajectory length) in **Figure 6.9(b)**.

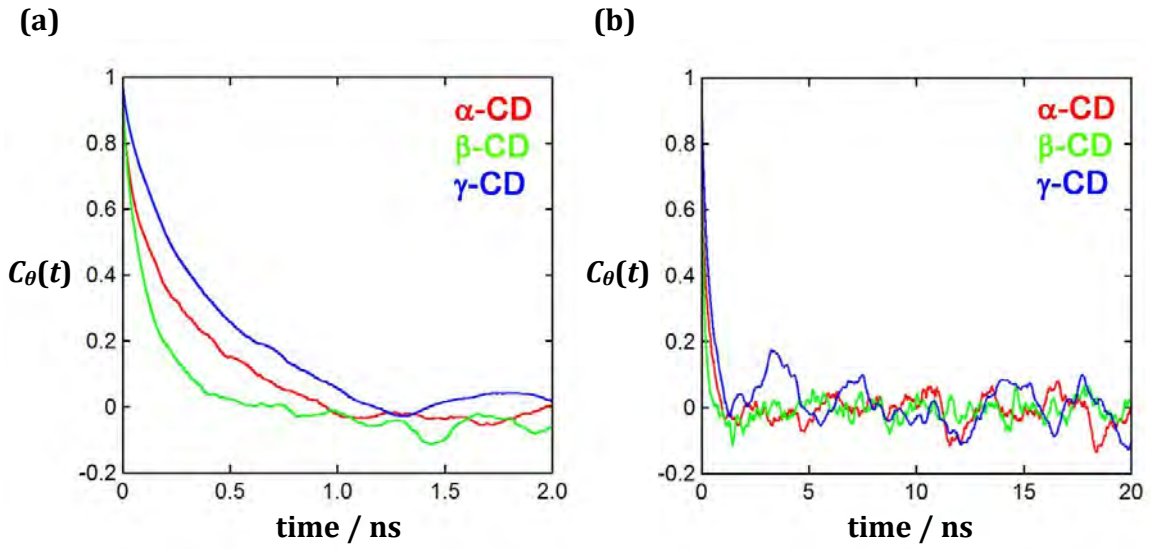


Figure 6.9: Average TCFs of the angular puckering coordinates of α-, β- and γ-CD, correlated over their 50 ns MD trajectories in water. In **(a)**, the TCFs are truncated at t = 2 ns, and in **(b)**, the TCFs are truncated at t = 20 ns.

Qualitatively, from **Figure 6.9** it can be seen that with respect to the puckering motion, β-CD relaxes the fastest, followed by α-CD and then γ-CD. To quantify the decay, correlation times were calculated for the TCFs by fitting them to the general form of the exponential decay function, given by Equation **6.4**:

$$y = A + B^{*}e^{(-t/\tau)} \qquad (6.4)$$

where A, B and τ are constants, and τ is the correlation time. The exponential fits were calculated using the first 1200 ps of data from the TCFs. The results of the fits are summarised in **Table 6.2**, where the τ values are reported with their associated standard errors in brackets, and the linear correlation coefficient (R) indicates the accuracy of the fit (the closer 'R' is to 1, the more accurate the fit).

Table 6.2: Results of fitting the average pucker TCFs for α-, β- and γ-CD to the general form of the exponential decay function (Equation **6.4**).

Cyclodextrin	τ (ps)	R
A	340.2 (± 0.7)	0.993
B	142.7 (± 0.2)	0.994
Γ	431.1 (± 0.6)	0.997

The order of the correlation times for the puckering motion is therefore $\tau_\gamma > \tau_\alpha > \tau_\beta$, quantitatively indicating that β-CD relaxes the fastest to the equilibrium structure, followed by α-CD and then γ-CD, where γ-CD is the slowest to reach equilibrium. The ratio of the τ values can also be calculated by dividing through by the smallest value (i.e. 142.7 ps). The ratio of the τ values (rounded to one decimal place) is as follows:

$$\tau_\alpha : \tau_\beta : \tau_\gamma \equiv 2.4 : 1 : 3.0$$

Thus β-CD decays approximately 2½ times faster than α-CD and three times faster than γ-CD. From the τ values and their relative ratio, it can thus be seen that with regards to the puckering motion of the macrocyclic ring, β-CD has the most restricted motion and is conformationally the least flexible, followed by α-CD and then γ-CD. This conclusion is consistent with analyses of the ψ glycosidic bond rotation and the monomeric motions in these CDs [3], the TCFs of which yielded the same relative rate of decay as that obtained for the macrocyclic puckering motion studied here. The order of CD flexibility for these three CDs studied here (as measured from the puckering motion) can also be correlated to their solubility trend, i.e. as the flexibility of the CD increases, so does its aqueous solubility – this observation agrees with that of previous studies carried out on these CDs [3, 5].

With regards to the probability distributions of **Figure 6.8**, although the ranges of θ show that β-CD has a slightly larger range for the puckering motion than α-CD, the *rate of decay* of the puckering (as obtained from the TCFs) is faster for β-CD than for α-CD, making β-CD conformationally the least flexible.

In order to explain the results of the pucker TCFs obtained here, there are different factors one can consider. Taking into account only the relative size of the CD rings,

smaller rings with fewer glucose units would be expected to have less freedom of movement on the macrocyclic scale, as smaller rings would be more restrained. Thus based on size alone, one would expect the order of ring flexibility to be γ-CD > β-CD > α-CD, but this is not the case (as β-CD is found to be less flexible than α-CD). Therefore there should be other factor/s contributing to the calculated order of flexibility. For example, one can consider the relative number of glucose units within the CD rings, where it is observed that α- and γ-CD (composed of an *even* number of glucose units) undergo macrocyclic puckering deformations more easily than β-CD (composed of an *odd* number of glucose units). Thus the lower flexibility of β-CD in comparison to the other two CDs could be investigated by considering the parity of the number of glucose residues in the rings. There is an increase in the relative flexibilities of α- and γ-CD as their ring size increases, therefore an increase in macrocyclic flexibility is intuitively expected as the number of monomers comprising the ring increases. There is no expectation for the flexibility to decrease with increase in ring size, as is observed here from α-CD to β-CD. As has been pointed out earlier, this unexpected microscopic observation corresponds with the unexpected macroscopic observation that there is a lowering of solubility from α-CD to β-CD.

In **Figure 6.9**, after the TCFs have decayed to zero, the correlation values fluctuate between positive and slightly negative values, giving rise to small-amplitude peaks in the TCFs. These fluctuations are more pronounced further along in the TCFs (see **Figure 6.9 (b)**), indicating the presence of other periodic motion/s in the system, as discussed in Chapter 3. These fluctuations can be analysed in future using Fourier analysis (as described in Chapter 3) to obtain information on the number and intensity of periodic motions present in the systems. In particular, another macrocyclic motion (in addition to the puckering) is evident from the MD trajectories – this is termed "breathing" in this work, and is defined and analysed in Section 6.4.2. The term "breathing" is used in vibrational spectroscopy in relation to bond deformations resulting from underlying changes to the electronic structure of small molecules. However, in this thesis, "breathing" refers to the *planar deformation motion of a macromolecule* – this concept is similar to the deformation motion studied by Zou *et al.* [33], who presented a method to describe the in-plane deformations of planar rings using a set of ring deformation

coordinates [33]. On the other hand, "breathing" in proteins (in aqueous solution) refers to the large-scale motions of the secondary structures, subunits or domains [34].

6.4.2 "Breathing" Motion

The "breathing" motion of the CDs is defined here as the planar deformation motion of the macrocyclic rings, i.e. the horizontal expansion and contraction (inward and outward deformations) of the macrocyclic rings. As the ring contracts inwards, it assumes a more oblong (or elliptical) shape, and as it expands outwards, it assumes a more symmetrical (or circular) shape. The "breathing" motion can be viewed as a dynamic equilibrium process, and is shown schematically in **Figure 6.10** for α-CD.

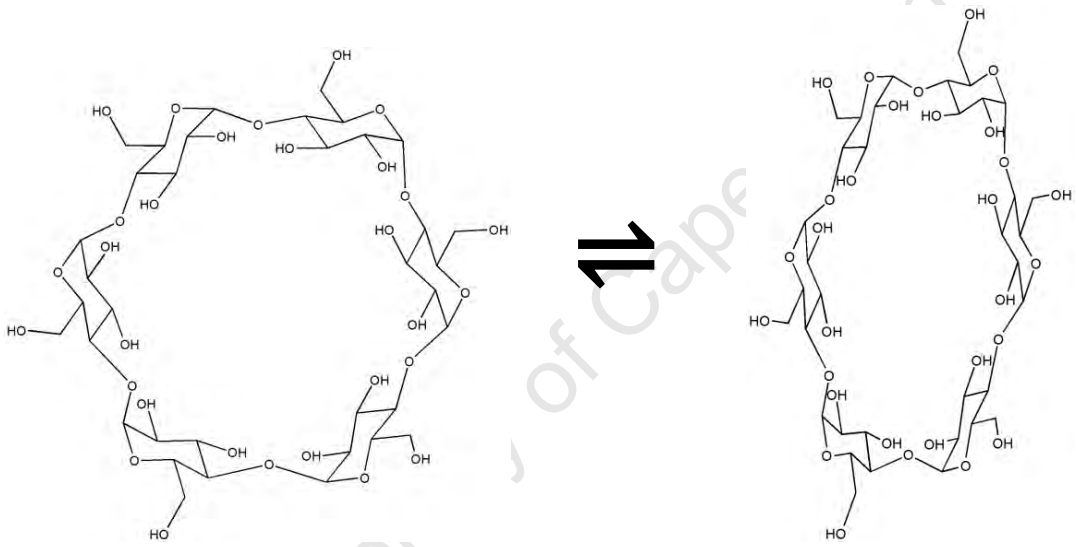


Figure 6.10: Schematic representation of the macrocyclic "breathing" motion in CDs, shown here for α-CD.

It is speculated that the "breathing" motion defined in this work could contribute to the fluctuations observed in the pucker TCFs. The coupling between the macrocyclic puckering motion and the "breathing" motion of the CDs is not explicitly investigated here, but may be interesting to investigate in the future. A systematic procedure is developed in this work to quantify the "breathing" motion observed in α-, β- and γ-CD. The method is presented for the general *N*-membered CD ring, using α-CD as an example to illustrate the points.

The *N*-membered CD rings are simplified in a coarse grain description (**Figure 6.7**) to polygons of size *N*, as done previously for the puckering analysis. This process is

illustrated in **Figure 6.11** for α-CD, where the conformation in **Figure 6.11(a)** is represented as a 6-sided 3D polygon in **Figure 6.11(b)** (which is drawn here in two dimensions for simplicity). The centres of mass of the macrocyclic rings are also calculated based on the dimer units' centres of mass (this is depicted by the black dot in **Figure 6.11**). **Figure 6.11** is drawn here symmetrically for illustrative purposes, and in reality, the conformations in the simulation do not have perfect symmetry. By analogy, β-CD and γ-CD are simplified to 7- and 8-sided 3D polygons, respectively. In **Figure 6.11(b)**, the polygon is divided into triangles to calculate the area-to-perimeter ratio of the polygon (as explained in the text following **Figure 6.11**).

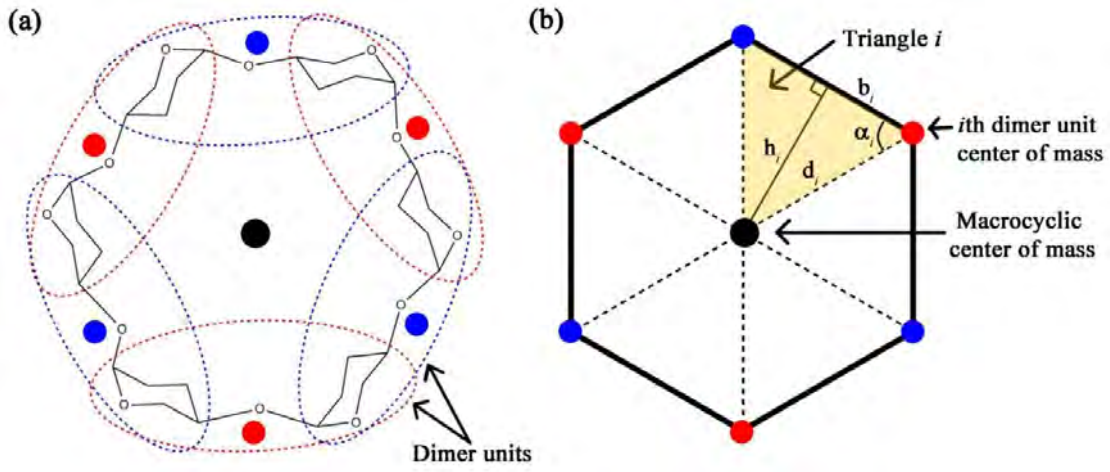


Figure 6.11: Schematic representation illustrating the quantification of the "breathing" motions for α-CD. On the left, **(a)** shows the dimer units (and their centres of mass as blue and red dots) of the macrocyclic ring of α-CD, which is reduced in a coarse grain representation to **(b)** a 3D polygon, which in the case of α-CD, is divided into six triangles to calculate the area-to-perimeter ratio of the polygon.

In order to quantify the "breathing" motions, the extent to which the 3D polygon is distorted from a circular shape will be measured. Here we define the term ***circularity*** as the degree to which the 3D polygon corresponding to the CD approximates a circle, i.e. how well the polygon's two-dimensional (2D) projection onto its mean plane fits into a circle. In this work however, a 2D projection of the polygon is not used quantitatively as the measure of circularity or the "breathing" motions, but rather the **area-to-perimeter** ratio of the polygon is compared to the **area-to-circumference** ratio of a reference circle.

This is done by dividing the polygon into *N* 2D triangles, where its area is defined as the sum of the areas of these triangles, and its perimeter is the sum of the 3D distances between the dimer units' centres of mass. The area-to-perimeter ratio of the polygon, termed $(\text{area / perimeter})_{CD}$, is then calculated as in Equation **6.5**.

$$(\text{area / perimeter})_{CD} = \frac{\sum_{i=1}^{N} \frac{1}{2} \mathbf{b}_i \mathbf{d}_i \sin \boldsymbol{\alpha}_i}{\sum_{i=1}^{N} \mathbf{b}_i} \qquad (\mathbf{6.5})$$

where $\mathbf{h}_i = \mathbf{d}_i \sin \boldsymbol{\alpha}_i$. For each triangle i ($1 \leq i \leq N$), $\mathbf{b}_i$ is the distance between adjacent dimer units' centres of mass, $\mathbf{d}_i$ is the distance between the macrocyclic centre of mass to the ith dimer unit centre of mass, and $\mathbf{h}_i$ is the perpendicular height of triangle i measured from the macrocyclic centre of mass to $\mathbf{b}_i$ – all distances are measured in three dimensions. $\boldsymbol{\alpha}_i$ is then the angle between $\mathbf{b}_i$ and $\mathbf{d}_i$ on the plane of triangle i.

The ratio $(\text{area / perimeter})_{CD}$ is then compared to the area-to-circumference ratio of a reference circle, to measure the extent to which the polygon is distorted from a circular shape. To define a suitable size for the reference circle, the standard deviation (SD) of $\mathbf{d}_i$ (**Figure 6.11(b)**) is calculated for every frame (i.e. conformation) in the 50 ns trajectories, as given by Equation **6.6**.

$$\text{SD} = \frac{\sqrt{\sum_{i=1}^{N} (\mathbf{d}_i - \mathbf{d}_{ave})^2}}{\sqrt{N-1}} \qquad (\mathbf{6.6})$$

where $\mathbf{d}_{ave}$ is the average of the N values of $\mathbf{d}_i$ for each conformation. The conformation with the lowest SD has the most circular shape in the trajectory, thus $\mathbf{d}_{ave}$ of this conformation, $\mathbf{d}'_{ave}$, is used to define the radius of a reference circle to which all the conformations will be compared. The area-to-circumference ratio of this circle, termed $(\text{area / circum})_{ref}$, is a constant defined as in Equation **6.7**.

$$(\text{area / circum})_{ref} = \frac{\mathbf{d}'_{ave}}{2} \qquad (\mathbf{6.7})$$

The ratios (area / perimeter)$_{CD}$ and (area / circum)$_{ref}$ are then compared to give the *breathing ratio*, as given in Equation **6.8**.

$$\text{breathing ratio} = \frac{(\text{area / perimeter})_{CD}}{(\text{area / circum})_{ref}} = \frac{\sum_{i=1}^{N} \mathbf{b}_i \mathbf{d}_i \sin \boldsymbol{\alpha}_i}{(\boldsymbol{d}'_{ave}) \times \sum_{i=1}^{N} \mathbf{b}_i} \qquad (\mathbf{6.8})$$

When the breathing ratio assumes values close to one, the polygon has a more symmetrical (or circular) shape, while values further from one indicate a more distorted (or oblong) shape. Thus the breathing ratio can be used as a quantitative measure of the "breathing" motion. This is explained by considering the ideal case where the polygon corresponding to the CD is 2D, as illustrated in **Figure 6.12** for α-CD.

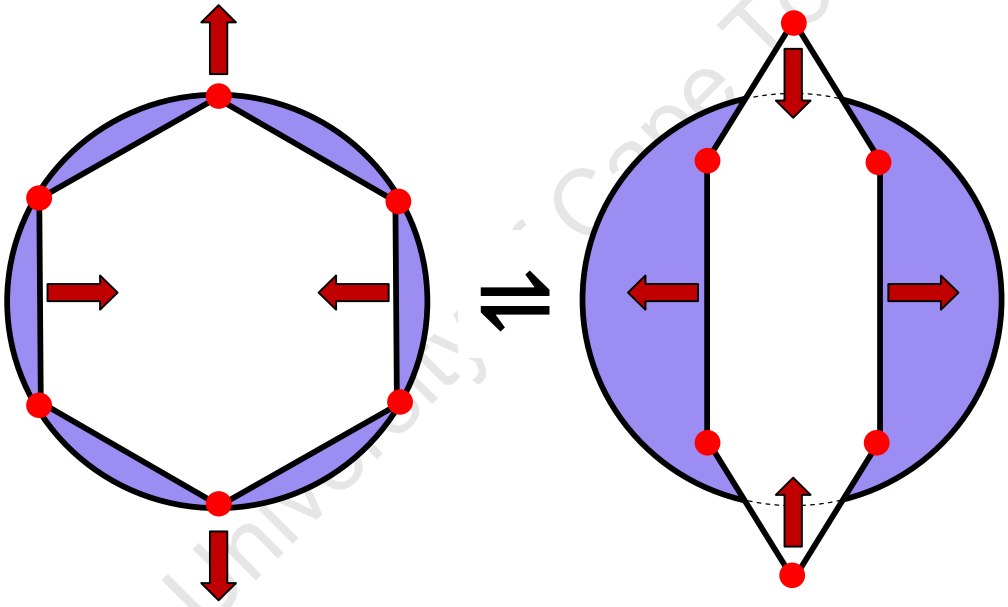


Figure 6.12: Schematic diagram of the "breathing" motion of a 2D polygon corresponding to the CD, shown here for a 6-membered polygon corresponding to α-CD. The polygon changes between a symmetrical shape on the left to a more oblong shape on the right, where the reference circle is shaded in blue. The red arrows represent the "breathing" motion of the ring.

In **Figure 6.12**, the 6-membered polygon on the left is in its most symmetrical form, i.e. a hexagon, which can be circumscribed in the reference circle (shaded in blue). The polygon on the right, however, is distorted to an oblong shape. The "breathing" of the CD ring (represented by the red arrows) causes the CD to change between the symmetrical and distorted forms. Qualitatively, it can be seen that the area and

perimeter of the symmetrical ring are close to that of the circle, thus (area / perimeter)$_{CD}$ is relatively close to (area / circum)$_{ref}$, and the breathing ratio is therefore close to one. As the symmetrical ring distorts to the oblong form, its area decreases while its perimeter changes relatively less, causing (area / perimeter)$_{CD}$ (and therefore the breathing ratio) to decrease for the oblong ring. Thus the breathing ratio can be considered as a measure of the "breathing" motion – as the breathing ratio tends towards one, the CD ring assumes a more circular shape, and as it decreases from one, the ring becomes more distorted. The above argument can be extended to the case of the 3D polygons associated with the CD rings here.

Time series of "breathing" motion

The breathing ratio was calculated for every frame in the CD trajectories and plotted as breathing time series (see **Figure 6.13(a)**). The time series indicate a *regular* "breathing" motion throughout all the CD trajectories, which possibly plays a role in the fluctuations observed in the pucker TCFs (**Figure 6.9**). Probability distributions of the breathing time series are given in **Figure 6.13(b)**, where the probabilities of the breathing ratio were calculated using intervals of 0.001 in the range (0.7, 1). The probabilities are plotted against the midpoint of the corresponding interval, e.g. the probability of the breathing ratio in the interval [0.8, 0.801] was plotted against breathing ratio = 0.8005.

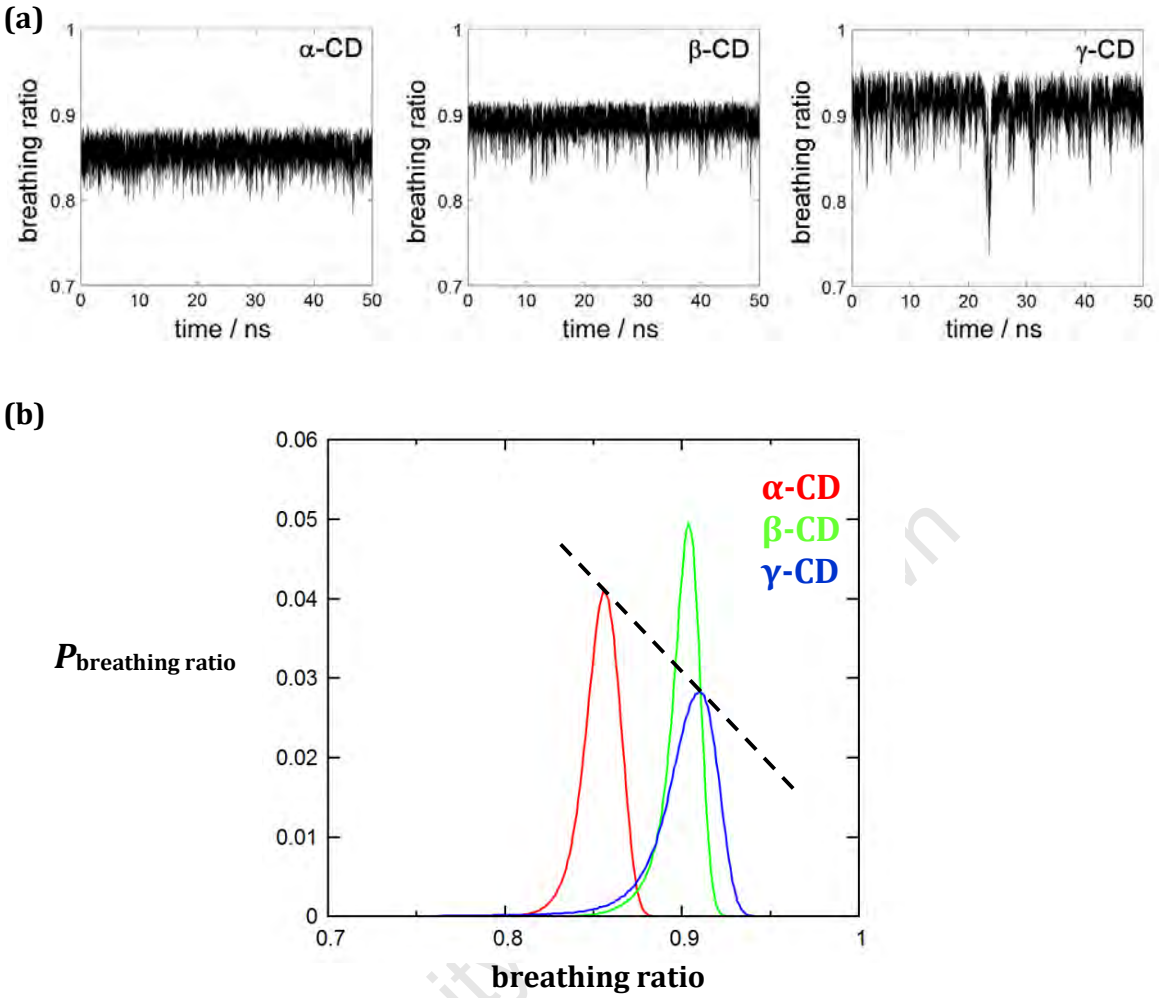


Figure 6.13: (a) Time series of the breathing ratios of α-, β- and γ-CDs over their 50 ns MD trajectories. **(b)** Probability distributions of the breathing ratios of α-, β- and γ-CDs. The dashed black line connecting the peak heights of α- and γ-CD indicates the relative order in the maximum probability heights of the CDs.

A significant point to note about the "breathing" distributions in **Figure 6.13(b)** is the ordering of the maximum probability heights (peak heights), which increase in the order γ-CD < α-CD < β-CD, as indicated by the dotted black line connecting the peak heights of α- and γ-CD. This order *correlates inversely* with the increasing CD solubility trend. A further observation is made by considering the ranges of the breathing ratio from **Figure 6.13(b)**, which are 0.107, 0.117 and 0.228 for α-, β- and γ-CD, respectively. This indicates an increase in the breathing ratio relative to the CD ring size, however, the increase in the range when going from α- to β-CD is 0.010, while the increase from β- to γ-CD is significantly larger, i.e. 0.111. The relatively narrow peak of β-CD and the order of the peak heights indicate that the "breathing" deformation is more limited in

this case. This implies that the more limited the in-plane "breathing" deformation motion is, the less soluble the CD is likely to be in water. This is reminiscent of the correlation between ring puckering and CD solubility. There is thus a possible coupling with the pucker derived flexibilities measured earlier and the in-plane ring deformation "breathing" motions.

The mean (±SD) of the distributions are 0.855 (±0.010), 0.891 (±0.010), and 0.911 (±0.020), and the maximum values of the breathing ratio are 0.889, 0.922 and 0.964 for α-, β- and γ-CD, respectively. Both the means and maximum values follow the order γ-CD > β-CD > α-CD, showing that α- and β-CD are less likely to assume a circular shape while the values for γ-CD reveal that this macrocycle approximates a circle best. This can be explained in terms of geometrical considerations, although this does not seem to be related to the solubility trend of the CDs. The definition of the breathing ratio for the 2D polygons corresponding to the CDs has the relation: *breathing ratio* = sin$\boldsymbol{\alpha}$, where the values of $\boldsymbol{\alpha}_i$ for each polygon are all equivalent. The breathing ratios are 0.866, 0.901, and 0.924 for the hexagon, heptagon and octagon, respectively, increasing in the order hexagon < heptagon < octagon. This shows that intrinsically, the octagon has the most circular shape, followed by the heptagon and then the hexagon. Therefore, for the corresponding CDs, γ-CD can attain the most circular shape, followed by β-CD and α-CD.

The "breathing" motion is illustrated in **Figure 6.14** using conformations of the CDs taken at different times from the 50 ns trajectories – the corresponding values of the breathing ratio (rounded to three decimal places) are also included. From **Figure 6.14**, it is shown that values of the breathing ratio close to one indicate a relatively circular shape for the CD, while those further from one correspond to a distorted shape.

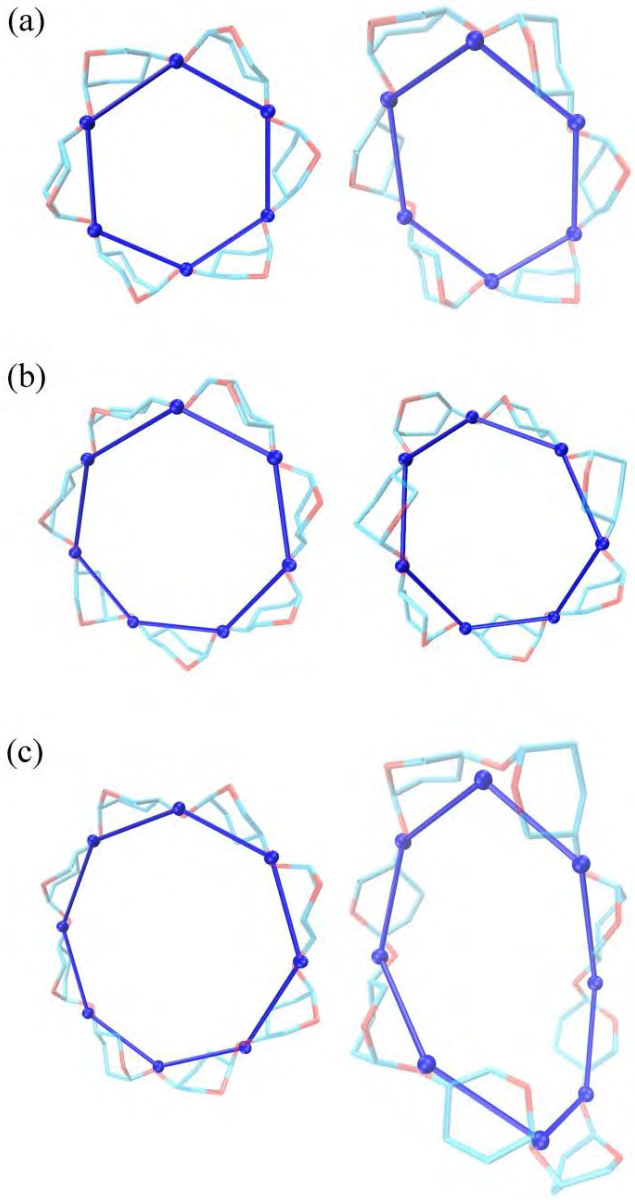


Figure 6.14: Illustration of the breathing motion in α-, β- and γ-CDs, using conformations taken at different times in their MD trajectories. **(a)** α-CD: breathing ratio = 0.882 (left) and 0.822 (right); **(b)** β-CD: breathing ratio = 0.912 (left) and 0.826 (right); and **(c)** γ-CD: breathing ratio = 0.942 (left) and 0.737 (right). 3D polygons based on the dimer units' centers of mass are shown (glucose ring substituents are omitted). Color code: Carbon = cyan, oxygen = red, centers of mass of dimer units = blue.

In summary, from the macrocyclic puckering analysis carried out in this work on α-, β- and γ-CD, the correlation between the relative flexibilities and the aqueous solubility trend of these three CDs has been established: as the puckering flexibility of the CD increases, so does its solubility. In addition, the "breathing" analysis showed an inverse

correlation with the solubility trend, i.e. as the maximum probability heights of the distributions of the breathing ratio increases, so the solubility of the CD decreases.

References

1. Dodziuk, H. *Journal of Molecular Structure* **2002**, *614*, 33.
2. Stezowski, J. J.; Parker, W.; Hilgenkamp, S.; Gdaniec, M. *Journal of the American Chemical Society* **2001**, *123*, 3919.
3. Naidoo, K. J.; Gamieldien, M. R.; Chen, J. Y.-J.; Widmalm, G. r.; Maliniak, A. *The Journal of Physical Chemistry B* **2008**, *112*, 15151.
4. Kozár, T.; Venanzi, C. A. *Journal of Molecular Structure: THEOCHEM* **1997**, *395–396*, 451.
5. Naidoo, K. J.; Chen, J. Y.-J.; Jansson, J. L. M.; Widmalm, G.; Maliniak, A. *The Journal of Physical Chemistry B* **2004**, *108*, 4236.
6. Pereira, C. S.; de Moura, A. F.; Freitas, L. C. G.; Lins, R. D. *Journal of the Brazilian Chemical Society* **2007**, *18*, 951.
7. Raffaini, G.; Ganazzoli, F. *Chemical Physics* **2007**, *333*, 128.
8. Ivanov, P. M. *The Journal of Physical Chemistry B* **2010**, *114*, 2650.
9. Gotsev, M. G.; Ivanov, P. M. *The Journal of Physical Chemistry B* **2009**, *113*, 5752.
10. Thaning, J.; Stevensson, B.; Östervall, J.; Naidoo, K. J.; Widmalm, G. r.; Maliniak, A. *The Journal of Physical Chemistry B* **2008**, *112*, 8434.
11. French, A. D.; Johnson, G. P. *Carbohydrate Research* **2007**, *342*, 1223.
12. Shikata, T.; Takahashi, R.; Satokawa, Y. *The Journal of Physical Chemistry B* **2007**, *111*, 12239.
13. Bernatowicz, P.; Ruszczyńska-Bartnik, K.; Ejchart, A.; Dodziuk, H.; Kaczorowska, E.; Ueda, H. *The Journal of Physical Chemistry B* **2010**, *114*, 59.
14. Perez-Miron, J.; Jaime, C.; Ivanov, P. M. *Chirality* **2008**, *20*, 1127.
15. Schneider, H.-J.; Hacket, F.; Rüdiger, V.; Ikeda, H. *Chemical Reviews* **1998**, *98*, 1755.
16. Manunza, B.; Deiana, S.; Pintore, M.; Gessa, C. *Journal of Molecular Structure: THEOCHEM* **1997**, *419*, 133.
17. Koehler, J. E. H.; Saenger, W.; van Gunsteren, W. F. *Journal of Molecular Biology* **1988**, *203*, 241.
18. Starikov, E. B.; Bräsicke, K.; Knapp, E. W.; Saenger, W. *Chemical Physics Letters* **2001**, *336*, 504.

19. Linert, W.; Margl, P.; Renz, F. *Chemical Physics* **1992**, *161*, 327.
20. Heine, T.; Dos Santos, H. F.; Patchkovskii, S.; Duarte, H. A. *The Journal of Physical Chemistry A* **2007**, *111*, 5648.
21. Rodriguez, J.; Hernán Rico, D.; Domenianni, L.; Laria, D. *The Journal of Physical Chemistry B* **2008**, *112*, 7522.
22. Jozwiakowski, M. J.; Connors, K. A. *Carbohydrate Research* **1985**, *143*, 51.
23. Kuttel, M.; Brady, J. W.; Naidoo, K. J. *Journal of Computational Chemistry* **2002**, *23*, 1236.
24. Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. *The Journal of Physical Chemistry* **1987**, *91*, 6269.
25. Guvench, O.; Greene, S. N.; Kamath, G.; Brady, J. W.; Venable, R. M.; Pastor, R. W.; Mackerell, A. D. *Journal of Computational Chemistry* **2008**, *29*, 2543.
26. Guvench, O.; Hatcher, E.; Venable, R. M.; Pastor, R. W.; MacKerell, A. D. *Journal of Chemical Theory and Computation* **2009**, *5*, 2353.
27. Guvench, O.; Mallajosyula, S. S.; Raman, E. P.; Hatcher, E.; Vanommeslaeghe, K.; Foster, T. J.; Jamison, F. W.; MacKerell, A. D. *Journal of Chemical Theory and Computation* **2011**, *7*, 3162.
28. Hatcher, E. R.; Guvench, O.; MacKerell, A. D. *Journal of Chemical Theory and Computation* **2009**, *5*, 1315.
29. Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *Journal of Computational Chemistry* **1983**, *4*, 187.
30. Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. *Journal of Computational Physics* **1977**, *23*, 327.
31. Feller, S. E.; Zhang, Y.; Pastor, R. W.; Brooks, B. R. *The Journal of Chemical Physics* **1995**, *103*, 4613.
32. Hockney, R. W. *Methods in Computational Physics* **1970**, *9*, 136.
33. Zou, W.; Izotov, D.; Cremer, D. *The Journal of Physical Chemistry A* **2011**, *115*, 8731.
34. Makowski, L.; Rodi, D. J.; Mandava, S.; Minh, D. D. L.; Gore, D. B.; Fischetti, R. F. *Journal of Molecular Biology* **2008**, *375*, 529.

Chapter 7

Conclusions

The objectives in this thesis were (i) to extend the triangular tessellation (TT) model for 6-membered monocyclic rings to medium-sized 7- and 8-membered rings, (ii) to develop a systematic procedure to generate the TT coordinates (angular puckering coordinates, or APC) of all the canonical states of an *N*-membered ring (*N* = 6, 7, 8), and (iii) to explain the anomalous aqueous solubility of β-CD, as compared to α- and γ-CD, by quantitatively analysing the macrocyclic ring puckering motions of these cyclodextrins (CDs) from MD simulations using the TT model. In particular, the puckering of the CDs was studied through a *coarse grained analysis* of the full atomistic MD simulations.

Objective (i) was met, by defining a system of triangular planes to define the ring puckering of 7- and 8-membered rings, and deriving mathematical expressions for the APC. Thus the description of ring conformations for the medium sized 7- and 8-membered rings can be successfully described using the TT model. A similar procedure used for the tessellation of these rings can be used in future work to describe the pucker of larger rings. However, for very large rings, the large number of APC (there are *N*-3 APC for an *N*-membered ring) may have a limited descriptive power for the ring pucker.

Objective (ii) was also met, and the systematic procedure was successfully used to generate the APC of the canonical states of cycloheptane and cyclooctane. This provides a useful "mapping" of a quantitative description of ring pucker (i.e. the APC) onto the qualitative description of the ring canonical conformations (i.e. nomenclature such as *chair*, *boat*, etc.) for 7- and 8-membered rings. This "mapping" can be particularly useful in future *dynamics* simulations, to monitor ring conformational changes quantitatively during the course of a trajectory and then translate them to qualitative descriptions that may be more meaningful to chemists. However, more work needs to be done on *systematically* assigning the qualitative descriptions during a dynamics simulation, where rings can take on several shapes over time that are not the *ideal* canonical shapes and do not correspond to the ideal APC.

Objective (iii) was also met, showing that APC can be used to analyse complex CD molecules, by using an appropriate coarse grained definition of the macrocyclic ring (i.e. simplifying the CD ring to a monomeric representation) – the method of TT can possibly be applied to other macrocyclic structures in future, to study their ring puckering. The relative CD flexibilities measured from time correlation functions of the APC follows the same order as the solubility trend, i.e. both decrease in the order γ-CD > α-CD > β-CD, showing that the conformational puckering flexibilities of these CD rings are related to their solubility trend. In addition, the maximum probability heights of the probability distributions of the CD breathing ratios follow the order β-CD > α-CD > γ-CD, which correlates inversely with the increasing CD solubility trend and indicates that the "breathing" deformation is more limited in the case of β-CD. Both the puckering and "breathing" analyses suggest that the anomalous solubility of β-CD is related to the macrocyclic conformational motions of the three CDs. It may also be useful in future to investigate the coupling between the macrocyclic puckering motion and the "breathing" motion of the CDs. Further investigation is being done on the correlation between the flexibility and solubility of these CDs following the work in this thesis.

In future work, the pucker TCFs of the CDs can be analysed by Fourier analysis to gain insight into the number and intensity of periodic motions present in these systems. This has not been carried out in this work, as the focus here was to identify the *nature* of the motions contributing to the CD ring flexibilities, and to provide insight into how the dynamic macrocyclic deformations observed for the three CDs are related to their solubility trend. Also, the idea of conformational flexibility can potentially be a useful parameter to explore in relation to the physical properties of other macrocyclic structures.

Appendix A

Table A1: Angular puckering coordinates of the 38 canonical conformations of cyclohexane, taken from reference [1] of Chapter 5. Conf. is an abbreviation for 'conformation'.

Conf.	θ_0 (°)	θ_1 (°)	θ_2 (°)
$^{1}C_{4}$	-35.26	-35.26	-35.26
$^{4}C_{1}$	35.26	35.26	35.26
$^{1,4}B$	-35.26	74.20	-35.26
$B_{1,4}$	35.26	-74.20	35.26
$^{2,5}B$	74.20	-35.26	-35.26
$B_{2,5}$	-74.20	35.26	35.26
$^{3,6}B$	-35.26	-35.26	74.20
$B_{3,6}$	35.26	35.26	-74.20
$^{1}H_{2}$	-42.16	9.07	-17.83
$^{2}H_{1}$	42.16	-9.07	17.83
$^{2}H_{3}$	42.16	17.83	-9.06
$^{3}H_{2}$	-42.16	-17.83	9.06
$^{3}H_{4}$	-17.83	-42.16	9.07
$^{4}H_{3}$	17.83	42.16	-9.07
$^{4}H_{5}$	-9.07	42.16	17.83
$^{5}H_{4}$	9.07	-42.16	-17.83
$^{5}H_{6}$	9.07	-17.83	-42.16
$^{6}H_{5}$	-9.07	17.83	42.16
$^{6}H_{1}$	17.83	-9.07	42.16
$^{1}H_{6}$	-17.83	9.07	-42.16

Conf.	θ_0 (°)	θ_1 (°)	θ_2 (°)
$^{1}S_{3}$	0	50.84	-50.84
$^{3}S_{1}$	0	-50.84	50.84
$^{5}S_{1}$	50.84	-50.84	0
$^{1}S_{5}$	-50.84	50.84	0
$^{6}S_{2}$	-50.84	0	50.84
$^{2}S_{6}$	50.84	0	-50.84
^{1}E	-35.26	17.37	-35.26
E_{1}	35.26	-17.37	35.26
^{2}E	46.86	0	0
E_{2}	-46.86	0	0
^{3}E	-35.26	-35.26	17.37
E_{3}	35.26	35.26	-17.37
^{4}E	0	46.86	0
E_{4}	0	-46.86	0
^{5}E	17.37	-35.26	-35.26
E_{5}	-17.37	35.26	35.26
^{6}E	0	0	46.86
E_{6}	0	0	-46.86

Table A2: Angular puckering coordinates of all the permutations and negations of the 38 canonical states of cyclohexane, generated by the permutation and negation cycles of Procedure 1 (Chapter 5).

chair

Conf.	θ_0 (°)	θ_1 (°)	θ_2 (°)	Set
$^1C_4$ – **P1**	-35.26	-35.26	-35.25	α
$^1C_4$ – **P2**	35.25	35.26	35.26	β
$^1C_4$ – **P3**	-35.26	-35.25	-35.26	α
$^1C_4$ – **P4**	35.26	35.26	35.25	β
$^1C_4$ – **P5**	-35.25	-35.26	-35.26	α
$^1C_4$ – **P6**	35.26	35.25	35.26	β
$^1C_4$ – **N1**	35.26	35.26	35.25	β
$^1C_4$ – **N2**	-35.25	-35.26	-35.26	α
$^1C_4$ – **N3**	35.26	35.25	35.26	β
$^1C_4$ – **N4**	-35.26	-35.26	-35.25	α
$^1C_4$ – **N5**	35.25	35.26	35.26	β
$^1C_4$ – **N6**	-35.26	-35.25	-35.26	α

Conf.	θ_0 (°)	θ_1 (°)	θ_2 (°)	Set
$^4C_1$ – **P1**	35.26	35.26	35.25	β
$^4C_1$ – **P2**	-35.25	-35.26	-35.26	α
$^4C_1$ – **P3**	35.26	35.25	35.26	β
$^4C_1$ – **P4**	-35.26	-35.26	-35.25	α
$^4C_1$ – **P5**	35.25	35.26	35.26	β
$^4C_1$ – **P6**	-35.26	-35.25	-35.26	α
$^4C_1$ – **N1**	-35.26	-35.26	-35.25	α
$^4C_1$ – **N2**	35.25	35.26	35.26	β
$^4C_1$ – **N3**	-35.26	-35.25	-35.26	α
$^4C_1$ – **N4**	35.26	35.26	35.25	β
$^4C_1$ – **N5**	-35.25	-35.26	-35.26	α
$^4C_1$ – **N6**	35.26	35.25	35.26	β

boat

Conf.	θ_0 (°)	θ_1 (°)	θ_2 (°)	Set
$^{1,4}B$ – **P1**	-35.26	74.20	-35.25	α
$^{1,4}B$ – **P2**	-35.25	-35.27	74.20	β
$^{1,4}B$ – **P3**	74.20	-35.25	-35.26	γ
$^{1,4}B$ – **P4**	-35.27	74.20	-35.25	α
$^{1,4}B$ – **P5**	-35.25	-35.26	74.20	β
$^{1,4}B$ – **P6**	74.20	-35.25	-35.27	γ
$^{1,4}B$ – **N1**	35.26	-74.20	35.25	–α
$^{1,4}B$ – **N2**	35.25	35.27	-74.20	–β
$^{1,4}B$ – **N3**	-74.20	35.25	35.26	–γ
$^{1,4}B$ – **N4**	35.27	-74.20	35.25	–α
$^{1,4}B$ – **N5**	35.25	35.26	-74.20	–β
$^{1,4}B$ – **N6**	-74.20	35.25	35.27	–γ

Conf.	θ_0 (°)	θ_1 (°)	θ_2 (°)	Set
$B_{1,4}$ – **P1**	35.26	-74.20	35.25	–α
$B_{1,4}$ – **P2**	35.25	35.27	-74.20	–β
$B_{1,4}$ – **P3**	-74.20	35.25	35.26	–γ
$B_{1,4}$ – **P4**	35.27	-74.20	35.25	–α
$B_{1,4}$ – **P5**	35.25	35.26	-74.20	–β
$B_{1,4}$ – **P6**	-74.20	35.25	35.27	–γ
$B_{1,4}$ – **N1**	-35.26	74.20	-35.25	α
$B_{1,4}$ – **N2**	-35.25	-35.27	74.20	β
$B_{1,4}$ – **N3**	74.20	-35.25	-35.26	γ
$B_{1,4}$ – **N4**	-35.27	74.20	-35.25	α
$B_{1,4}$ – **N5**	-35.25	-35.26	74.20	β
$B_{1,4}$ – **N6**	74.20	-35.25	-35.27	γ

$^{2,5}B$ – **P1**	74.20	-35.27	-35.26	γ	$B_{2,5}$ – **P1**	-74.20	35.27	35.26	–γ
$^{2,5}B$ – **P2**	-35.25	74.21	-35.26	α	$B_{2,5}$ – **P2**	35.25	-74.21	35.26	–α
$^{2,5}B$ – **P3**	-35.27	-35.26	74.20	β	$B_{2,5}$ – **P3**	35.27	35.26	-74.20	–β
$^{2,5}B$ – **P4**	74.21	-35.26	-35.25	γ	$B_{2,5}$ – **P4**	-74.21	35.26	35.25	–γ
$^{2,5}B$ – **P5**	-35.26	74.20	-35.27	α	$B_{2,5}$ – **P5**	35.26	-74.20	35.27	–α
$^{2,5}B$ – **P6**	-35.26	-35.25	74.21	β	$B_{2,5}$ – **P6**	35.26	35.25	-74.21	–β
$^{2,5}B$ – **N1**	-74.20	35.27	35.26	–γ	$B_{2,5}$ – **N1**	74.20	-35.27	-35.26	γ
$^{2,5}B$ – **N2**	35.25	-74.21	35.26	–α	$B_{2,5}$ – **N2**	-35.25	74.21	-35.26	α
$^{2,5}B$ – **N3**	35.27	35.26	-74.20	–β	$B_{2,5}$ – **N3**	-35.27	-35.26	74.20	β
$^{2,5}B$ – **N4**	-74.21	35.26	35.25	–γ	$B_{2,5}$ – **N4**	74.21	-35.26	-35.25	γ
$^{2,5}B$ – **N5**	35.26	-74.20	35.27	–α	$B_{2,5}$ – **N5**	-35.26	74.20	-35.27	α
$^{2,5}B$ – **N6**	35.26	35.25	-74.21	–β	$B_{2,5}$ – **N6**	-35.26	-35.25	74.21	β
$^{3,6}B$ – **P1**	-35.26	-35.26	74.18	β	$B_{3,6}$ – **P1**	35.26	35.26	-74.18	–β
$^{3,6}B$ – **P2**	74.18	-35.26	-35.26	γ	$B_{3,6}$ – **P2**	-74.18	35.26	35.26	–γ
$^{3,6}B$ – **P3**	-35.26	74.18	-35.26	α	$B_{3,6}$ – **P3**	35.26	-74.18	35.26	–α
$^{3,6}B$ – **P4**	-35.26	-35.26	74.18	β	$B_{3,6}$ – **P4**	35.26	35.26	-74.18	–β
$^{3,6}B$ – **P5**	74.18	-35.26	-35.26	γ	$B_{3,6}$ – **P5**	-74.18	35.26	35.26	–γ
$^{3,6}B$ – **P6**	-35.26	74.18	-35.26	α	$B_{3,6}$ – **P6**	35.26	-74.18	35.26	–α
$^{3,6}B$ – **N1**	35.26	35.26	-74.18	–β	$B_{3,6}$ – **N1**	-35.26	-35.26	74.18	β
$^{3,6}B$ – **N2**	-74.18	35.26	35.26	–γ	$B_{3,6}$ – **N2**	74.18	-35.26	-35.26	γ
$^{3,6}B$ – **N3**	35.26	-74.18	35.26	–α	$B_{3,6}$ – **N3**	-35.26	74.18	-35.26	α
$^{3,6}B$ – **N4**	35.26	35.26	-74.18	–β	$B_{3,6}$ – **N4**	-35.26	-35.26	74.18	β
$^{3,6}B$ – **N5**	-74.18	35.26	35.26	–γ	$B_{3,6}$ – **N5**	74.18	-35.26	-35.26	γ
$^{3,6}B$ – **N6**	35.26	-74.18	35.26	–α	$B_{3,6}$ – **N6**	-35.26	74.18	-35.26	α

half-chair

Conf.	θ_0 (°)	θ_1 (°)	θ_2 (°)	Set	Conf.	θ_0 (°)	θ_1 (°)	θ_2 (°)	Set
$^{1}H_2$ – **P1**	-42.16	9.04	-17.83	α	$^{2}H_1$ – **P1**	42.16	-9.04	17.83	–α
$^{1}H_2$ – **P2**	15.61	-7.89	40.22	β	$^{2}H_1$ – **P2**	-15.61	7.89	-40.22	–β
$^{1}H_2$ – **P3**	9.04	-17.83	-42.16	γ	$^{2}H_1$ – **P3**	-9.04	17.83	42.16	–γ
$^{1}H_2$ – **P4**	-7.89	40.22	15.61	δ	$^{2}H_1$ – **P4**	7.89	-40.22	-15.61	–δ
$^{1}H_2$ – **P5**	-17.83	-42.16	9.04	ε	$^{2}H_1$ – **P5**	17.83	42.16	-9.04	–ε
$^{1}H_2$ – **P6**	40.22	15.61	-7.89	ζ	$^{2}H_1$ – **P6**	-40.22	-15.61	7.89	–ζ
$^{1}H_2$ – **N1**	42.16	-9.04	17.83	–α	$^{2}H_1$ – **N1**	-42.16	9.04	-17.83	α
$^{1}H_2$ – **N2**	-15.61	7.89	-40.22	–β	$^{2}H_1$ – **N2**	15.61	-7.89	40.22	β
$^{1}H_2$ – **N3**	-9.04	17.83	42.16	–γ	$^{2}H_1$ – **N3**	9.04	-17.83	-42.16	γ
$^{1}H_2$ – **N4**	7.89	-40.22	-15.61	–δ	$^{2}H_1$ – **N4**	-7.89	40.22	15.61	δ
$^{1}H_2$ – **N5**	17.83	42.16	-9.04	–ε	$^{2}H_1$ – **N5**	-17.83	-42.16	9.04	ε
$^{1}H_2$ – **N6**	-40.22	-15.61	7.89	–ζ	$^{2}H_1$ – **N6**	40.22	15.61	-7.89	ζ

$^{2}H_3$ – **P1**	42.16	17.83	-9.06	ζ	$^{3}H_2$ – **P1**	-42.16	-17.83	9.06	–ζ
$^{2}H_3$ – **P2**	-40.23	7.90	-15.61	α	$^{3}H_2$ – **P2**	40.23	-7.90	15.61	–α
$^{2}H_3$ – **P3**	17.83	-9.06	42.16	β	$^{3}H_2$ – **P3**	-17.83	9.06	-42.16	–β
$^{2}H_3$ – **P4**	7.90	-15.61	-40.23	γ	$^{3}H_2$ – **P4**	-7.90	15.61	40.23	–γ
$^{2}H_3$ – **P5**	-9.06	42.16	17.83	δ	$^{3}H_2$ – **P5**	9.06	-42.16	-17.83	–δ
$^{2}H_3$ – **P6**	-15.61	-40.23	7.90	ε	$^{3}H_2$ – **P6**	15.61	40.23	-7.90	–ε
$^{2}H_3$ – **N1**	-42.16	-17.83	9.06	–ζ	$^{3}H_2$ – **N1**	42.16	17.83	-9.06	ζ
$^{2}H_3$ – **N2**	40.23	-7.90	15.61	–α	$^{3}H_2$ – **N2**	-40.23	7.90	-15.61	α
$^{2}H_3$ – **N3**	-17.83	9.06	-42.16	–β	$^{3}H_2$ – **N3**	17.83	-9.06	42.16	β
$^{2}H_3$ – **N4**	-7.90	15.61	40.23	–γ	$^{3}H_2$ – **N4**	7.90	-15.61	-40.23	γ
$^{2}H_3$ – **N5**	9.06	-42.16	-17.83	–δ	$^{3}H_2$ – **N5**	-9.06	42.16	17.83	δ
$^{2}H_3$ – **N6**	15.61	40.23	-7.90	–ε	$^{3}H_2$ – **N6**	-15.61	-40.23	7.90	ε
$^{3}H_4$ – **P1**	-17.83	-42.17	9.10	ε	$^{4}H_3$ – **P1**	17.83	42.17	-9.10	–ε
$^{3}H_4$ – **P2**	40.26	15.58	-7.93	ζ	$^{4}H_3$ – **P2**	-40.26	-15.58	7.93	–ζ
$^{3}H_4$ – **P3**	-42.17	9.10	-17.83	α	$^{4}H_3$ – **P3**	42.17	-9.10	17.83	–α
$^{3}H_4$ – **P4**	15.58	-7.93	40.26	β	$^{4}H_3$ – **P4**	-15.58	7.93	-40.26	–β
$^{3}H_4$ – **P5**	9.10	-17.83	-42.17	γ	$^{4}H_3$ – **P5**	-9.10	17.83	42.17	–γ
$^{3}H_4$ – **P6**	-7.93	40.26	15.58	δ	$^{4}H_3$ – **P6**	7.93	-40.26	-15.58	–δ
$^{3}H_4$ – **N1**	17.83	42.17	-9.10	–ε	$^{4}H_3$ – **N1**	-17.83	-42.17	9.10	ε
$^{3}H_4$ – **N2**	-40.26	-15.58	7.93	–ζ	$^{4}H_3$ – **N2**	40.26	15.58	-7.93	ζ
$^{3}H_4$ – **N3**	42.17	-9.10	17.83	–α	$^{4}H_3$ – **N3**	-42.17	9.10	-17.83	α
$^{3}H_4$ – **N4**	-15.58	7.93	-40.26	–β	$^{4}H_3$ – **N4**	15.58	-7.93	40.26	β
$^{3}H_4$ – **N5**	-9.10	17.83	42.17	–γ	$^{4}H_3$ – **N5**	9.10	-17.83	-42.17	γ
$^{3}H_4$ – **N6**	7.93	-40.26	-15.58	–δ	$^{4}H_3$ – **N6**	-7.93	40.26	15.58	δ
$^{4}H_5$ – **P1**	-9.08	42.17	17.86	δ	$^{5}H_4$ – **P1**	9.08	-42.17	-17.86	–δ
$^{4}H_5$ – **P2**	-15.59	-40.23	7.90	ε	$^{5}H_4$ – **P2**	15.59	40.23	-7.90	–ε
$^{4}H_5$ – **P3**	42.17	17.86	-9.08	ζ	$^{5}H_4$ – **P3**	-42.17	-17.86	9.08	–ζ
$^{4}H_5$ – **P4**	-40.23	7.90	-15.59	α	$^{5}H_4$ – **P4**	40.23	-7.90	15.59	–α
$^{4}H_5$ – **P5**	17.86	-9.08	42.17	β	$^{5}H_4$ – **P5**	-17.86	9.08	-42.17	–β
$^{4}H_5$ – **P6**	7.90	-15.59	-40.23	γ	$^{5}H_4$ – **P6**	-7.90	15.59	40.23	–γ
$^{4}H_5$ – **N1**	9.08	-42.17	-17.86	–δ	$^{5}H_4$ – **N1**	-9.08	42.17	17.86	δ
$^{4}H_5$ – **N2**	15.59	40.23	-7.90	–ε	$^{5}H_4$ – **N2**	-15.59	-40.23	7.90	ε
$^{4}H_5$ – **N3**	-42.17	-17.86	9.08	–ζ	$^{5}H_4$ – **N3**	42.17	17.86	-9.08	ζ
$^{4}H_5$ – **N4**	40.23	-7.90	15.59	–α	$^{5}H_4$ – **N4**	-40.23	7.90	-15.59	α
$^{4}H_5$ – **N5**	-17.86	9.08	-42.17	–β	$^{5}H_4$ – **N5**	17.86	-9.08	42.17	β
$^{4}H_5$ – **N6**	-7.90	15.59	40.23	–γ	$^{5}H_4$ – **N6**	7.90	-15.59	-40.23	γ

$^{5}H_{6}$ – **P1**	9.08	-17.84	-42.15	γ
$^{5}H_{6}$ – **P2**	-7.90	40.25	15.58	δ
$^{5}H_{6}$ – **P3**	-17.84	-42.15	9.08	ε
$^{5}H_{6}$ – **P4**	40.25	15.58	-7.90	ζ
$^{5}H_{6}$ – **P5**	-42.15	9.08	-17.84	α
$^{5}H_{6}$ – **P6**	15.58	-7.90	40.25	β
$^{5}H_{6}$ – **N1**	-9.08	17.84	42.15	–γ
$^{5}H_{6}$ – **N2**	7.90	-40.25	-15.58	–δ
$^{5}H_{6}$ – **N3**	17.84	42.15	-9.08	–ε
$^{5}H_{6}$ – **N4**	-40.25	-15.58	7.90	–ζ
$^{5}H_{6}$ – **N5**	42.15	-9.08	17.84	–α
$^{5}H_{6}$ – **N6**	-15.58	7.90	-40.25	–β

$^{6}H_{5}$ – **P1**	-9.08	17.84	42.15	–γ
$^{6}H_{5}$ – **P2**	7.90	-40.25	-15.58	–δ
$^{6}H_{5}$ – **P3**	17.84	42.15	-9.08	–ε
$^{6}H_{5}$ – **P4**	-40.25	-15.58	7.90	–ζ
$^{6}H_{5}$ – **P5**	42.15	-9.08	17.84	–α
$^{6}H_{5}$ – **P6**	-15.58	7.90	-40.25	–β
$^{6}H_{5}$ – **N1**	9.08	-17.84	-42.15	γ
$^{6}H_{5}$ – **N2**	-7.90	40.25	15.58	δ
$^{6}H_{5}$ – **N3**	-17.84	-42.15	9.08	ε
$^{6}H_{5}$ – **N4**	40.25	15.58	-7.90	ζ
$^{6}H_{5}$ – **N5**	-42.15	9.08	-17.84	α
$^{6}H_{5}$ – **N6**	15.58	-7.90	40.25	β

$^{6}H_{1}$ – **P1**	17.83	-9.09	42.15	β
$^{6}H_{1}$ – **P2**	7.92	-15.59	-40.25	γ
$^{6}H_{1}$ – **P3**	-9.09	42.15	17.83	δ
$^{6}H_{1}$ – **P4**	-15.59	-40.25	7.92	ε
$^{6}H_{1}$ – **P5**	42.15	17.83	-9.09	ζ
$^{6}H_{1}$ – **P6**	-40.25	7.92	-15.59	α
$^{6}H_{1}$ – **N1**	-17.83	9.09	-42.15	–β
$^{6}H_{1}$ – **N2**	-7.92	15.59	40.25	–γ
$^{6}H_{1}$ – **N3**	9.09	-42.15	-17.83	–δ
$^{6}H_{1}$ – **N4**	15.59	40.25	-7.92	–ε
$^{6}H_{1}$ – **N5**	-42.15	-17.83	9.09	–ζ
$^{6}H_{1}$ – **N6**	40.25	-7.92	15.59	–α

$^{1}H_{6}$ – **P1**	-17.83	9.09	-42.15	–β
$^{1}H_{6}$ – **P2**	-7.92	15.59	40.25	–γ
$^{1}H_{6}$ – **P3**	9.09	-42.15	-17.83	–δ
$^{1}H_{6}$ – **P4**	15.59	40.25	-7.92	–ε
$^{1}H_{6}$ – **P5**	-42.15	-17.83	9.09	–ζ
$^{1}H_{6}$ – **P6**	40.25	-7.92	15.59	–α
$^{1}H_{2}$ – **N1**	17.83	-9.09	42.15	β
$^{1}H_{6}$ – **N2**	7.92	-15.59	-40.25	γ
$^{1}H_{6}$ – **N3**	-9.09	42.15	17.83	δ
$^{1}H_{6}$ – **N4**	-15.59	-40.25	7.92	ε
$^{1}H_{6}$ – **N5**	42.15	17.83	-9.09	ζ
$^{1}H_{6}$ – **N6**	-40.25	7.92	-15.59	α

twist-boat

Conf.	θ_0 (°)	θ_1 (°)	θ_2 (°)	Set
$^{1}S_{3}$ – **P1**	0	50.82	-50.84	α
$^{1}S_{3}$ – **P2**	-50.84	0.02	50.83	β
$^{1}S_{3}$ – **P3**	50.82	-50.84	0	γ
$^{1}S_{3}$ – **P4**	0.02	50.83	-50.84	α
$^{1}S_{3}$ – **P5**	-50.84	0	50.82	β
$^{1}S_{3}$ – **P6**	50.83	-50.84	0.02	γ
$^{1}S_{3}$ – **N1**	0	-50.82	50.84	–α
$^{1}S_{3}$ – **N2**	50.84	-0.02	-50.83	–β
$^{1}S_{3}$ – **N3**	-50.82	50.84	0	–γ
$^{1}S_{3}$ – **N4**	-0.02	-50.83	50.84	–α
$^{1}S_{3}$ – **N5**	50.84	0	-50.82	–β
$^{1}S_{3}$ – **N6**	-50.83	50.84	-0.02	–γ

Conf.	θ_0 (°)	θ_1 (°)	θ_2 (°)	Set
$^{3}S_{1}$ – **P1**	0	-50.82	50.84	–α
$^{3}S_{1}$ – **P2**	50.84	-0.02	-50.83	–β
$^{3}S_{1}$ – **P3**	-50.82	50.84	0	–γ
$^{3}S_{1}$ – **P4**	-0.02	-50.83	50.84	–α
$^{3}S_{1}$ – **P5**	50.84	0	-50.82	–β
$^{3}S_{1}$ – **P6**	-50.83	50.84	-0.02	–γ
$^{3}S_{1}$ – **N1**	0	50.82	-50.84	α
$^{3}S_{1}$ – **N2**	-50.84	0.02	50.83	β
$^{3}S_{1}$ – **N3**	50.82	-50.84	0	γ
$^{3}S_{1}$ – **N4**	0.02	50.83	-50.84	α
$^{3}S_{1}$ – **N5**	-50.84	0	50.82	β
$^{3}S_{1}$ – **N6**	50.83	-50.84	0.02	γ

$^1S_5$ – **P1**	-50.85	50.85	0.01	–γ
$^1S_5$ – **P2**	0	-50.84	50.84	–α
$^1S_5$ – **P3**	50.85	0.01	-50.85	–β
$^1S_5$ – **P4**	-50.84	50.84	0	–γ
$^1S_5$ – **P5**	0.01	-50.85	50.85	–α
$^1S_5$ – **P6**	50.84	0	-50.84	–β
$^1S_5$ – **N1**	50.85	-50.85	-0.01	γ
$^1S_5$ – **N2**	0	50.84	-50.84	α
$^1S_5$ – **N3**	-50.85	-0.01	50.85	β
$^1S_5$ – **N4**	50.84	-50.84	0	γ
$^1S_5$ – **N5**	-0.01	50.85	-50.85	α
$^1S_5$ – **N6**	-50.84	0	50.84	β

$^5S_1$ – **P1**	50.85	-50.85	-0.01	γ
$^5S_1$ – **P2**	0	50.84	-50.84	α
$^5S_1$ – **P3**	-50.85	-0.01	50.85	β
$^5S_1$ – **P4**	50.84	-50.84	0	γ
$^5S_1$ – **P5**	-0.01	50.85	-50.85	α
$^5S_1$ – **P6**	-50.84	0	50.84	β
$^5S_1$ – **N1**	-50.85	50.85	0.01	–γ
$^5S_1$ – **N2**	0	-50.84	50.84	–α
$^5S_1$ – **N3**	50.85	0.01	-50.85	–β
$^5S_1$ – **N4**	-50.84	50.84	0	–γ
$^5S_1$ – **N5**	0.01	-50.85	50.85	–α
$^5S_1$ – **N6**	50.84	0	-50.84	–β

$^2S_6$ – **P1**	50.85	-0.05	-50.88	–β
$^2S_6$ – **P2**	-50.83	50.90	0	–γ
$^2S_6$ – **P3**	-0.05	-50.88	50.85	–α
$^2S_6$ – **P4**	50.90	0	-50.83	–β
$^2S_6$ – **P5**	-50.88	50.85	-0.05	–γ
$^2S_6$ – **P6**	0	-50.83	50.90	–α
$^2S_6$ – **N1**	-50.85	0.05	50.88	β
$^2S_6$ – **N2**	50.83	-50.90	0	γ
$^2S_6$ – **N3**	0.05	50.88	-50.85	α
$^2S_6$ – **N4**	-50.90	0	50.83	β
$^2S_6$ – **N5**	50.88	-50.85	0.05	γ
$^2S_6$ – **N6**	0	50.83	-50.90	α

$^6S_2$ – **P1**	-50.85	0.05	50.88	β
$^6S_2$ – **P2**	50.83	-50.90	0	γ
$^6S_2$ – **P3**	0.05	50.88	-50.85	α
$^6S_2$ – **P4**	-50.90	0	50.83	β
$^6S_2$ – **P5**	50.88	-50.85	0.05	γ
$^6S_2$ – **P6**	0	50.83	-50.90	α
$^6S_2$ – **N1**	50.85	-0.05	-50.88	–β
$^6S_2$ – **N2**	-50.83	50.90	0	–γ
$^6S_2$ – **N3**	-0.05	-50.88	50.85	–α
$^6S_2$ – **N4**	50.90	0	-50.83	–β
$^6S_2$ – **N5**	-50.88	50.85	-0.05	–γ
$^6S_2$ – **N6**	0	-50.83	50.90	–α

envelope

Conf.	θ_0 (°)	θ_1 (°)	θ_2 (°)	Set
^1E – **P1**	-35.27	17.38	-35.28	α
^1E – **P2**	0.43	0.44	48.8	β
^1E – **P3**	17.38	-35.28	-35.27	γ
^1E – **P4**	0.44	48.8	0.43	δ
^1E – **P5**	-35.28	-35.27	17.38	ε
^1E – **P6**	48.8	0.43	0.44	ζ
^1E – **N1**	35.27	-17.38	35.28	–α
^1E – **N2**	-0.43	-0.44	-48.8	–β
^1E – **N3**	-17.38	35.28	35.27	–γ
^1E – **N4**	-0.44	-48.8	-0.43	–δ
^1E – **N5**	35.28	35.27	-17.38	–ε
^1E – **N6**	-48.8	-0.43	-0.44	–ζ

Conf.	θ_0 (°)	θ_1 (°)	θ_2 (°)	Set
E_1 – **P1**	35.27	-17.38	35.28	–α
E_1 – **P2**	-0.43	-0.44	-48.8	–β
E_1 – **P3**	-17.38	35.28	35.27	–γ
E_1 – **P4**	-0.44	-48.8	-0.43	–δ
E_1 – **P5**	35.28	35.27	-17.38	–ε
E_1 – **P6**	-48.8	-0.43	-0.44	–ζ
E_1 – **N1**	-35.27	17.38	-35.28	α
E_1 – **N2**	0.43	0.44	48.8	β
E_1 – **N3**	17.38	-35.28	-35.27	γ
E_1 – **N4**	0.44	48.8	0.43	δ
E_1 – **N5**	-35.28	-35.27	17.38	ε
E_1 – **N6**	48.8	0.43	0.44	ζ

^2E – **P1**	46.85	0.05	0.02	ζ	E_2 – **P1**	-46.85	-0.05	-0.02	–ζ
^2E – **P2**	-29.43	15.74	-29.43	α	E_2 – **P2**	29.43	-15.74	29.43	–α
^2E – **P3**	0.05	0.02	46.85	β	E_2 – **P3**	-0.05	-0.02	-46.85	–β
^2E – **P4**	15.74	-29.43	-29.43	γ	E_2 – **P4**	-15.74	29.43	29.43	–γ
^2E – **P5**	0.02	46.85	0.05	δ	E_2 – **P5**	-0.02	-46.85	-0.05	–δ
^2E – **P6**	-29.43	-29.43	15.74	ε	E_2 – **P6**	29.43	29.43	-15.74	–ε
^2E – **N1**	-46.85	-0.05	-0.02	–ζ	E_2 – **N1**	46.85	0.05	0.02	ζ
^2E – **N2**	29.43	-15.74	29.43	–α	E_2 – **N2**	-29.43	15.74	-29.43	α
^2E – **N3**	-0.05	-0.02	-46.85	–β	E_2 – **N3**	0.05	0.02	46.85	β
^2E – **N4**	-15.74	29.43	29.43	–γ	E_2 – **N4**	15.74	-29.43	-29.43	γ
^2E – **N5**	-0.02	-46.85	-0.05	–δ	E_2 – **N5**	0.02	46.85	0.05	δ
^2E – **N6**	29.43	29.43	-15.74	–ε	E_2 – **N6**	-29.43	-29.43	15.74	ε
^3E – **P1**	-35.27	-35.21	17.41	ε	E_3 – **P1**	35.27	35.21	-17.41	–ε
^3E – **P2**	48.8	0.38	0.43	ζ	E_3 – **P2**	-48.8	-0.38	-0.43	–ζ
^3E – **P3**	-35.21	17.41	-35.27	α	E_3 – **P3**	35.21	-17.41	35.27	–α
^3E – **P4**	0.38	0.43	48.8	β	E_3 – **P4**	-0.38	-0.43	-48.8	–β
^3E – **P5**	17.41	-35.27	-35.21	γ	E_3 – **P5**	-17.41	35.27	35.21	–γ
^3E – **P6**	0.43	48.8	0.38	δ	E_3 – **P6**	-0.43	-48.8	-0.38	–δ
^3E – **N1**	35.27	35.21	-17.41	–ε	E_3 – **N1**	-35.27	-35.21	17.41	ε
^3E – **N2**	-48.8	-0.38	-0.43	–ζ	E_3 – **N2**	48.8	0.38	0.43	ζ
^3E – **N3**	35.21	-17.41	35.27	–α	E_3 – **N3**	-35.21	17.41	-35.27	α
^3E – **N4**	-0.38	-0.43	-48.8	–β	E_3 – **N4**	0.38	0.43	48.8	β
^3E – **N5**	-17.41	35.27	35.21	–γ	E_3 – **N5**	17.41	-35.27	-35.21	γ
^3E – **N6**	-0.43	-48.8	-0.38	–δ	E_3 – **N6**	0.43	48.8	0.38	δ
^4E – **P1**	0	46.87	0	δ	E_4 – **P1**	0	-46.87	0	–δ
^4E – **P2**	-29.41	-29.41	15.78	ε	E_4 – **P2**	29.41	29.41	-15.78	–ε
^4E – **P3**	46.87	0	0	ζ	E_4 – **P3**	-46.87	0	0	–ζ
^4E – **P4**	-29.41	15.78	-29.41	α	E_4 – **P4**	29.41	-15.78	29.41	–α
^4E – **P5**	0	0	46.87	β	E_4 – **P5**	0	0	-46.87	–β
^4E – **P6**	15.78	-29.41	-29.41	γ	E_4 – **P6**	-15.78	29.41	29.41	–γ
^4E – **N1**	0	-46.87	0	–δ	E_4 – **N1**	0	46.87	0	δ
^4E – **N2**	29.41	29.41	-15.78	–ε	E_4 – **N2**	-29.41	-29.41	15.78	ε
^4E – **N3**	-46.87	0	0	–ζ	E_4 – **N3**	46.87	0	0	ζ
^4E – **N4**	29.41	-15.78	29.41	–α	E_4 – **N4**	-29.41	15.78	-29.41	α
^4E – **N5**	0	0	-46.87	–β	E_4 – **N5**	0	0	46.87	β
^4E – **N6**	-15.78	29.41	29.41	–γ	E_4 – **N6**	15.78	-29.41	-29.41	γ

^5E – **P1**	17.38	-35.26	-35.23	γ	E_5 – **P1**	-17.38	35.26	35.23	–γ
^5E – **P2**	0.43	48.83	0.43	δ	E_5 – **P2**	-0.43	-48.83	-0.43	–δ
^5E – **P3**	-35.26	-35.23	17.38	ε	E_5 – **P3**	35.26	35.23	-17.38	–ε
^5E – **P4**	48.83	0.43	0.43	ζ	E_5 – **P4**	-48.83	-0.43	-0.43	–ζ
^5E – **P5**	-35.23	17.38	-35.26	α	E_5 – **P5**	35.23	-17.38	35.26	–α
^5E – **P6**	0.43	0.43	48.83	β	E_5 – **P6**	-0.43	-0.43	-48.83	–β
^5E – **N1**	-17.38	35.26	35.23	–γ	E_5 – **N1**	17.38	-35.26	-35.23	γ
^5E – **N2**	-0.43	-48.83	-0.43	–δ	E_5 – **N2**	0.43	48.83	0.43	δ
^5E – **N3**	35.26	35.23	-17.38	–ε	E_5 – **N3**	-35.26	-35.23	17.38	ε
^5E – **N4**	-48.83	-0.43	-0.43	–ζ	E_5 – **N4**	48.83	0.43	0.43	ζ
^5E – **N5**	35.23	-17.38	35.26	–α	E_5 – **N5**	-35.23	17.38	-35.26	α
^5E – **N6**	-0.43	-0.43	-48.83	–β	E_5 – **N6**	0.43	0.43	48.83	β
^6E – **P1**	0	0	46.83	β	E_6 – **P1**	0	0	-46.83	–β
^6E – **P2**	15.78	-29.42	-29.41	γ	E_6 – **P2**	-15.78	29.42	29.41	–γ
^6E – **P3**	0	46.83	0	δ	E_6 – **P3**	0	-46.83	0	–δ
^6E – **P4**	-29.42	-29.41	15.78	ε	E_6 – **P4**	29.42	29.41	-15.78	–ε
^6E – **P5**	46.83	0	0	ζ	E_6 – **P5**	-46.83	0	0	–ζ
^6E – **P6**	-29.41	15.78	-29.42	α	E_6 – **P6**	29.41	-15.78	29.42	–α
^6E – **N1**	0	0	-46.83	–β	E_6 – **N1**	0	0	46.83	β
^6E – **N2**	-15.78	29.42	29.41	–γ	E_6 – **N2**	15.78	-29.42	-29.41	γ
^6E – **N3**	0	-46.83	0	–δ	E_6 – **N3**	0	46.83	0	δ
^6E – **N4**	29.42	29.41	-15.78	–ε	E_6 – **N4**	-29.42	-29.41	15.78	ε
^6E – **N5**	-46.83	0	0	–ζ	E_6 – **N5**	46.83	0	0	ζ
^6E – **N6**	29.41	-15.78	29.42	–α	E_6 – **N6**	-29.41	15.78	-29.42	α

Table A3: Final angular puckering coordinates of the permutations and negations of the 38 states of cyclohexane, after duplicate puckering coordinates have been removed (Chapter 5).

chair

Conf.	θ_0 (°)	θ_1 (°)	θ_2 (°)	Set
$^1C_4$ – **P1**	-35.26	-35.26	-35.25	α
$^1C_4$ – **P2**	35.25	35.26	35.26	β
$^4C_1$ – **P1**	35.26	35.26	35.25	β
$^4C_1$ – **P2**	-35.25	-35.26	-35.26	α

boat

Conf.	θ_0 (°)	θ_1 (°)	θ_2 (°)	Set
$^{1,4}B$ – **P1**	-35.26	74.20	-35.25	α
$^{1,4}B$ – **P2**	-35.25	-35.27	74.20	β
$^{1,4}B$ – **P3**	74.20	-35.25	-35.26	γ
$^{1,4}B$ – **N1**	35.26	-74.20	35.25	–α
$^{1,4}B$ – **N2**	35.25	35.27	-74.20	–β
$^{1,4}B$ – **N3**	-74.20	35.25	35.26	–γ
$^{2,5}B$ – **P1**	74.20	-35.27	-35.26	γ
$^{2,5}B$ – **P2**	-35.25	74.21	-35.26	α
$^{2,5}B$ – **P3**	-35.27	-35.26	74.20	β
$^{2,5}B$ – **N1**	-74.20	35.27	35.26	–γ
$^{2,5}B$ – **N2**	35.25	-74.21	35.26	–α
$^{2,5}B$ – **N3**	35.27	35.26	-74.20	–β
$^{3,6}B$ – **P1**	-35.26	-35.26	74.18	β
$^{3,6}B$ – **P2**	74.18	-35.26	-35.26	γ
$^{3,6}B$ – **P3**	-35.26	74.18	-35.26	α
$^{3,6}B$ – **N1**	35.26	35.26	-74.18	–β
$^{3,6}B$ – **N2**	-74.18	35.26	35.26	–γ
$^{3,6}B$ – **N3**	35.26	-74.18	35.26	–α

Conf.	θ_0 (°)	θ_1 (°)	θ_2 (°)	Set
$B_{1,4}$ – **P1**	35.26	-74.20	35.25	–α
$B_{1,4}$ – **P2**	35.25	35.27	-74.20	–β
$B_{1,4}$ – **P3**	-74.20	35.25	35.26	–γ
$B_{1,4}$ – **N1**	-35.26	74.20	-35.25	α
$B_{1,4}$ – **N2**	-35.25	-35.27	74.20	β
$B_{1,4}$ – **N3**	74.20	-35.25	-35.26	γ
$B_{2,5}$ – **P1**	-74.20	35.27	35.26	–γ
$B_{2,5}$ – **P2**	35.25	-74.21	35.26	–α
$B_{2,5}$ – **P3**	35.27	35.26	-74.20	–β
$B_{2,5}$ – **N1**	74.20	-35.27	-35.26	γ
$B_{2,5}$ – **N2**	-35.25	74.21	-35.26	α
$B_{2,5}$ – **N3**	-35.27	-35.26	74.20	β
$B_{3,6}$ – **P1**	35.26	35.26	-74.18	–β
$B_{3,6}$ – **P2**	-74.18	35.26	35.26	–γ
$B_{3,6}$ – **P3**	35.26	-74.18	35.26	–α
$B_{3,6}$ – **N1**	-35.26	-35.26	74.18	β
$B_{3,6}$ – **N2**	74.18	-35.26	-35.26	γ
$B_{3,6}$ – **N3**	-35.26	74.18	-35.26	α

twist-boat

Conf.	θ_0 (°)	θ_1 (°)	θ_2 (°)	Set
$^1S_3$ – **P1**	0	50.82	-50.84	α
$^1S_3$ – **P2**	-50.84	0.02	50.83	β
$^1S_3$ – **P3**	50.82	-50.84	0	γ
$^1S_3$ – **N1**	0	-50.82	50.84	–α
$^1S_3$ – **N2**	50.84	-0.02	-50.83	–β
$^1S_3$ – **N3**	-50.82	50.84	0	–γ
$^1S_5$ – **P1**	-50.85	50.85	0.01	–γ
$^1S_5$ – **P2**	0	-50.84	50.84	–α
$^1S_5$ – **P3**	50.85	0.01	-50.85	–β
$^1S_5$ – **N1**	50.85	-50.85	-0.01	γ
$^1S_5$ – **N2**	0	50.84	-50.84	α
$^1S_5$ – **N3**	-50.85	-0.01	50.85	β
$^2S_6$ – **P1**	50.85	-0.05	-50.88	–β
$^2S_6$ – **P2**	-50.83	50.90	0	–γ
$^2S_6$ – **P3**	-0.05	-50.88	50.85	–α
$^2S_6$ – **N1**	-50.85	0.05	50.88	β
$^2S_6$ – **N2**	50.83	-50.90	0	γ
$^2S_6$ – **N3**	0.05	50.88	-50.85	α

Conf.	θ_0 (°)	θ_1 (°)	θ_2 (°)	Set
$^3S_1$ – **P1**	0	-50.82	50.84	–α
$^3S_1$ – **P2**	50.84	-0.02	-50.83	–β
$^3S_1$ – **P3**	-50.82	50.84	0	–γ
$^3S_1$ – **N1**	0	50.82	-50.84	α
$^3S_1$ – **N2**	-50.84	0.02	50.83	β
$^3S_1$ – **N3**	50.82	-50.84	0	γ
$^5S_1$ – **P1**	50.85	-50.85	-0.01	γ
$^5S_1$ – **P2**	0	50.84	-50.84	α
$^5S_1$ – **P3**	-50.85	-0.01	50.85	β
$^5S_1$ – **N1**	-50.85	50.85	0.01	–γ
$^5S_1$ – **N2**	0	-50.84	50.84	–α
$^5S_1$ – **N3**	50.85	0.01	-50.85	–β
$^6S_2$ – **P1**	-50.85	0.05	50.88	β
$^6S_2$ – **P2**	50.83	-50.90	0	γ
$^6S_2$ – **P3**	0.05	50.88	-50.85	α
$^6S_2$ – **N1**	50.85	-0.05	-50.88	–β
$^6S_2$ – **N2**	-50.83	50.90	0	–γ
$^6S_2$ – **N3**	-0.05	-50.88	50.85	–α

half-chair & *envelope*

No duplicate states were generated from the permutation and negation cycles of the *half-chair* and *envelope*. Thus the final angular puckering coordinates of these states are the same as those given in **Table A2**.

Appendix B

Table B1: Angular puckering coordinates of all the states of the five canonical shapes of cycloheptane, generated by Procedure 1 (Chapter 5).

boat

Conformer	θ_0 (°)	θ_1 (°)	θ_2 (°)	θ_3 (°)
boat – **P1**	-77.50	32.40	-49.77	65.42
boat – **P2**	32.40	-77.50	50.32	35.47
boat – **P3**	60.84	1.17	29.27	-78.25
boat – **P4**	-60.66	78.70	-60.66	0.00
boat – **P5**	-32.67	-32.67	0.00	77.12
boat – **P6**	78.70	-60.66	61.49	-37.04
boat – **P7**	1.17	60.84	-29.12	-60.82
boat – **N1**	77.50	-32.40	49.77	-65.42
boat – **N2**	-32.40	77.50	-50.32	-35.47
boat – **N3**	-60.84	-1.17	-29.27	78.25
boat – **N4**	60.66	-78.70	60.66	0.00
boat – **N5**	32.67	32.67	0.00	-77.12
boat – **N6**	-78.70	60.66	-61.49	37.04
boat – **N7**	-1.17	-60.84	29.12	60.82

chair

Conformer	θ_0 (°)	θ_1 (°)	θ_2 (°)	θ_3 (°)
chair – **P1**	-54.36	-30.98	70.78	0.65
chair – **P2**	71.29	-30.69	-53.87	-0.60
chair – **P3**	-30.69	71.29	54.20	-33.13
chair – **P4**	-30.98	-54.36	-70.65	44.72
chair – **P5**	55.76	27.58	55.76	0.00
chair – **P6**	-35.45	-35.45	0.00	-38.42
chair – **P7**	27.58	55.76	-57.33	33.96
chair – **N1**	54.36	30.98	-70.78	-0.65
chair – **N2**	-71.29	30.69	53.87	0.60
chair – **N3**	30.69	-71.29	-54.20	33.13
chair – **N4**	30.98	54.36	70.65	-44.72
chair – **N5**	-55.76	-27.58	-55.76	0.00
chair – **N6**	35.45	35.45	0.00	38.42
chair – **N7**	-27.58	-55.76	57.33	-33.96

twist-boat

Conformer	θ_0 (°)	θ_1 (°)	θ_2 (°)	θ_3 (°)
TB – **P1**	-17.13	-48.16	14.82	70.64
TB – **P2**	80.69	-47.04	57.54	-53.15
TB – **P3**	-16.09	70.42	-41.09	-49.32
TB – **P4**	-70.42	16.09	-40.66	73.73
TB – **P5**	47.04	-80.69	57.15	18.55
TB – **P6**	48.16	17.13	14.98	-79.53
TB – **P7**	-71.86	71.86	-61.83	18.76
TB – **N1**	17.13	48.16	-14.82	-70.64
TB – **N2**	-80.69	47.04	-57.54	53.15
TB – **N3**	16.09	-70.42	41.09	49.32
TB – **N4**	70.42	-16.09	40.66	-73.73
TB – **N5**	-47.04	80.69	-57.15	-18.55
TB – **N6**	-48.16	-17.13	-14.98	79.53
TB – **N7**	71.86	-71.86	61.83	-18.76

twist-chair

Conformer	θ_0 (°)	θ_1 (°)	θ_2 (°)	θ_3 (°)
TC – **P1**	56.52	34.60	73.14	-31.51
TC – **P2**	-48.50	-21.04	-34.76	-27.07
TC – **P3**	21.04	48.50	-34.82	47.68
TC – **P4**	-34.60	-56.52	74.95	-15.03
TC – **P5**	71.20	3.43	-59.71	-14.57
TC – **P6**	-63.16	63.16	45.49	-13.53
TC – **P7**	-3.43	-71.20	-61.43	51.46
TC – **N1**	-56.52	-34.60	-73.14	31.51
TC – **N2**	48.50	21.04	34.76	27.07
TC – **N3**	-21.04	-48.50	34.82	-47.68
TC – **N4**	34.60	56.52	-74.95	15.03
TC – **N5**	-71.20	-3.43	59.71	14.57
TC – **N6**	63.16	-63.16	-45.49	13.53
TC – **N7**	3.43	71.20	61.43	-51.46

TS3

Conformer	θ_0 (°)	θ_1 (°)	θ_2 (°)	θ_3 (°)
TS3 – **P1**	-67.64	-13.42	21.49	58.38
TS3 – **P2**	75.86	-75.86	13.92	-4.02
TS3 – **P3**	13.42	67.64	22.74	-71.28
TS3 – **P4**	-77.84	10.76	-75.34	50.69
TS3 – **P5**	36.70	-32.43	42.72	43.69
TS3 – **P6**	32.43	-36.70	43.21	-69.63
TS3 – **P7**	-10.76	77.84	-72.25	-4.45
TS3 – **N1**	67.64	13.42	-21.49	-58.38
TS3 – **N2**	-75.86	75.86	-13.92	4.02
TS3 – **N3**	-13.42	-67.64	-22.74	71.28
TS3 – **N4**	77.84	-10.76	75.34	-50.69
TS3 – **N5**	-36.70	32.43	-42.72	-43.69
TS3 – **N6**	-32.43	36.70	-43.21	69.63
TS3 – **N7**	10.76	-77.84	72.25	4.45

Appendix C

Table C1: Angular puckering coordinates of all the states of the eleven canonical shapes of cyclooctane, generated by Procedure 1 (Chapter 5).

boat-boat

Conformer	θ_0 (°)	θ_1 (°)	θ_2 (°)	θ_3 (°)	θ_4 (°)
BB – **P1**	-29.79	-29.79	-29.79	-29.79	89.26
BB – **P2**	74.02	-74.02	74.02	-74.02	0.00
BB – **N1**	29.79	29.79	29.79	29.79	-89.26
BB – **N2**	-74.02	74.02	-74.02	74.02	0.00

boat-chair

Conformer	θ_0 (°)	θ_1 (°)	θ_2 (°)	θ_3 (°)	θ_4 (°)
BC – **P1**	25.13	70.07	-86.70	70.07	0.00
BC – **P2**	-80.28	-7.61	-7.61	-80.28	67.51
BC – **P3**	70.07	-86.70	70.07	25.13	0.00
BC – **P4**	36.93	36.93	-35.62	-35.62	-62.16
BC – **P5**	-86.70	70.07	25.13	70.07	0.00
BC – **P6**	-7.61	-80.28	-80.28	-7.61	67.51
BC – **P7**	70.07	25.13	70.07	-86.70	0.00
BC – **P8**	-35.62	-35.62	36.93	36.93	-62.16
BC – **N1**	-25.13	-70.07	86.70	-70.07	0.00
BC – **N2**	80.28	7.61	7.61	80.28	-67.51
BC – **N3**	-70.07	86.70	-70.07	-25.13	0.00
BC – **N4**	-36.93	-36.93	35.62	35.62	62.16
BC – **N5**	86.70	-70.07	-25.13	-70.07	0.00
BC – **N6**	7.61	80.28	80.28	7.61	-67.51
BC – **N7**	-70.07	-25.13	-70.07	86.70	0.00
BC – **N8**	35.62	35.62	-36.93	-36.93	62.16

boat

Conformer	θ_0 (°)	θ_1 (°)	θ_2 (°)	θ_3 (°)	θ_4 (°)
boat – **P1**	-30.73	78.11	-30.73	78.11	-69.26
boat – **P2**	-78.11	30.73	-78.11	30.73	69.26
boat – **N1**	30.73	-78.11	30.73	-78.11	69.26
boat – **N2**	78.11	-30.73	78.11	-30.73	-69.26

chair

Conformer	θ_0 (°)	θ_1 (°)	θ_2 (°)	θ_3 (°)	θ_4 (°)
chair – **P1**	79.68	31.92	-79.68	-31.92	0.00
chair – **P2**	-79.68	31.92	79.68	-31.92	0.00
chair – **P3**	31.92	-79.68	-31.92	79.68	0.00
chair – **P4**	31.92	79.68	-31.92	-79.68	0.00
chair – **N1**	-79.68	-31.92	79.68	31.92	0.00
chair – **N2**	79.68	-31.92	-79.68	31.92	0.00
chair – **N3**	-31.92	79.68	31.92	-79.68	0.00
chair – **N4**	-31.92	-79.68	31.92	79.68	0.00

crown

Conformer	θ_0 (°)	θ_1 (°)	θ_2 (°)	θ_3 (°)	θ_4 (°)
Crown – **P1**	46.44	46.44	46.44	46.44	0.00
Crown – **N1**	-46.44	-46.44	-46.44	-46.44	0.00

twist boat-chair

Conformer	θ_0 (°)	θ_1 (°)	θ_2 (°)	θ_3 (°)	θ_4 (°)
TBC – **P1**	84.52	9.50	-44.91	-51.54	-34.52
TBC – **P2**	-72.03	74.58	67.75	13.82	-31.37
TBC – **P3**	-13.82	-67.75	-74.58	72.03	31.37
TBC – **P4**	51.54	44.91	-9.50	-84.52	34.52
TBC – **P5**	-44.91	-51.54	84.52	9.50	-34.52
TBC – **P6**	67.75	13.82	-72.03	74.58	-31.37
TBC – **P7**	-74.58	72.03	-13.82	-67.75	31.37
TBC – **P8**	-9.50	-84.52	51.54	44.91	34.52
TBC – **N1**	-84.52	-9.50	44.91	51.54	34.52
TBC – **N2**	72.03	-74.58	-67.75	-13.82	31.37
TBC – **N3**	13.82	67.75	74.58	-72.03	-31.37
TBC – **N4**	-51.54	-44.91	9.50	84.52	-34.52
TBC – **N5**	44.91	51.54	-84.52	-9.50	34.52
TBC – **N6**	-67.75	-13.82	72.03	-74.58	31.37
TBC – **N7**	74.58	-72.03	13.82	67.75	-31.37
TBC – **N8**	9.50	84.52	-51.54	-44.91	-34.52

twist chair-chair

Conformer	θ_0 (°)	θ_1 (°)	θ_2 (°)	θ_3 (°)	θ_4 (°)
TCC – **P1**	38.66	68.40	38.66	68.40	-19.82
TCC – **P2**	-68.40	-38.66	-68.40	-38.66	19.82
TCC – **P3**	55.39	25.68	55.39	25.68	16.89
TCC – **P4**	-25.68	-55.39	-25.68	-55.39	-16.89
TCC – **N1**	-38.66	-68.40	-38.66	-68.40	19.82
TCC – **N2**	68.40	38.66	68.40	38.66	-19.82
TCC – **N3**	-55.39	-25.68	-55.39	-25.68	-16.89
TCC – **N4**	25.68	55.39	25.68	55.39	16.89

TS1

Conformer	θ_0 (°)	θ_1 (°)	θ_2 (°)	θ_3 (°)	θ_4 (°)
TS1 – **P1**	7.35	-82.58	-34.62	-79.39	55.31
TS1 – **P2**	79.39	34.62	82.58	-7.35	-55.31
TS1 – **P3**	-62.32	0.86	-45.33	42.64	-44.65
TS1 – **P4**	-0.86	62.32	-42.64	45.33	44.65
TS1 – **P5**	-34.62	-79.39	7.35	-82.58	55.31
TS1 – **P6**	82.58	-7.35	79.39	34.62	-55.31
TS1 – **P7**	-45.33	42.64	-62.32	0.86	-44.65
TS1 – **P8**	-42.64	45.33	-0.86	62.32	44.65
TS1 – **N1**	-7.35	82.58	34.62	79.39	-55.31
TS1 – **N2**	-79.39	-34.62	-82.58	7.35	55.31
TS1 – **N3**	62.32	-0.86	45.33	-42.64	44.65
TS1 – **N4**	0.86	-62.32	42.64	-45.33	-44.65
TS1 – **N5**	34.62	79.39	-7.35	82.58	-55.31
TS1 – **N6**	-82.58	7.35	-79.39	-34.62	55.31
TS1 – **N7**	45.33	-42.64	62.32	-0.86	44.65
TS1 – **N8**	42.64	-45.33	0.86	-62.32	-44.65

TS2

Conformer	θ_0 (°)	θ_1 (°)	θ_2 (°)	θ_3 (°)	θ_4 (°)
TS2 – **P1**	-85.65	24.00	30.46	67.92	22.44
TS2 – **P2**	39.58	-81.16	-73.88	-19.82	52.53
TS2 – **P3**	39.42	45.72	83.90	-78.70	-21.74
TS2 – **P4**	-45.57	-37.39	16.57	75.46	-51.38
TS2 – **P5**	30.46	67.92	-85.65	24.00	22.44
TS2 – **P6**	-73.88	-19.82	39.58	-81.16	52.53
TS2 – **P7**	83.90	-78.70	39.42	45.72	-21.74
TS2 – **P8**	16.57	75.46	-45.57	-37.39	-51.38
TS2 – **N1**	85.65	-24.00	-30.46	-67.92	-22.44
TS2 – **N2**	-39.58	81.16	73.88	19.82	-52.53
TS2 – **N3**	-39.42	-45.72	-83.90	78.70	21.74
TS2 – **N4**	45.57	37.39	-16.57	-75.46	51.38
TS2 – **N5**	-30.46	-67.92	85.65	-24.00	-22.44
TS2 – **N6**	73.88	19.82	-39.58	81.16	-52.53
TS2 – **N7**	-83.90	78.70	-39.42	-45.72	21.74
TS2 – **N8**	-16.57	-75.46	45.57	37.39	51.38

TS3

Conformer	θ_0 (°)	θ_1 (°)	θ_2 (°)	θ_3 (°)	θ_4 (°)
TS3 – **P1**	60.43	-60.43	-60.43	60.43	0.00
TS3 – **P2**	0.00	88.06	0.00	-88.06	0.00
TS3 – **P3**	-60.43	-60.43	60.43	60.43	0.00
TS3 – **P4**	88.06	0.00	-88.06	0.00	0.00
TS3 – **N1**	-60.43	60.43	60.43	-60.43	0.00
TS3 – **N2**	0.00	-88.06	0.00	88.06	0.00
TS3 – **N3**	60.43	60.43	-60.43	-60.43	0.00
TS3 – **N4**	-88.06	0.00	88.06	0.00	0.00

TS4

Conformer	θ_0 (°)	θ_1 (°)	θ_2 (°)	θ_3 (°)	θ_4 (°)
TS4 – **P1**	-20.14	-56.34	-56.34	-20.14	85.60
TS4 – **P2**	75.80	-30.96	75.80	-88.87	0.00
TS4 – **P3**	-0.67	-0.67	37.69	37.69	-83.25
TS4 – **P4**	-30.96	75.80	-88.87	75.80	0.00
TS4 – **P5**	-56.34	-20.14	-20.14	-56.34	85.60
TS4 – **P6**	75.80	-88.87	75.80	-30.96	0.00
TS4 – **P7**	37.69	37.69	-0.67	-0.67	-83.25
TS4 – **P8**	-88.87	75.80	-30.96	75.80	0.00
TS4 – **N1**	20.14	56.34	56.34	20.14	-85.60
TS4 – **N2**	-75.80	30.96	-75.80	88.87	0.00
TS4 – **N3**	0.67	0.67	-37.69	-37.69	83.25
TS4 – **N4**	30.96	-75.80	88.87	-75.80	0.00
TS4 – **N5**	56.34	20.14	20.14	56.34	-85.60
TS4 – **N6**	-75.80	88.87	-75.80	30.96	0.00
TS4 – **N7**	-37.69	-37.69	0.67	0.67	83.25
TS4 – **N8**	88.87	-75.80	30.96	-75.80	0.00

Appendix D

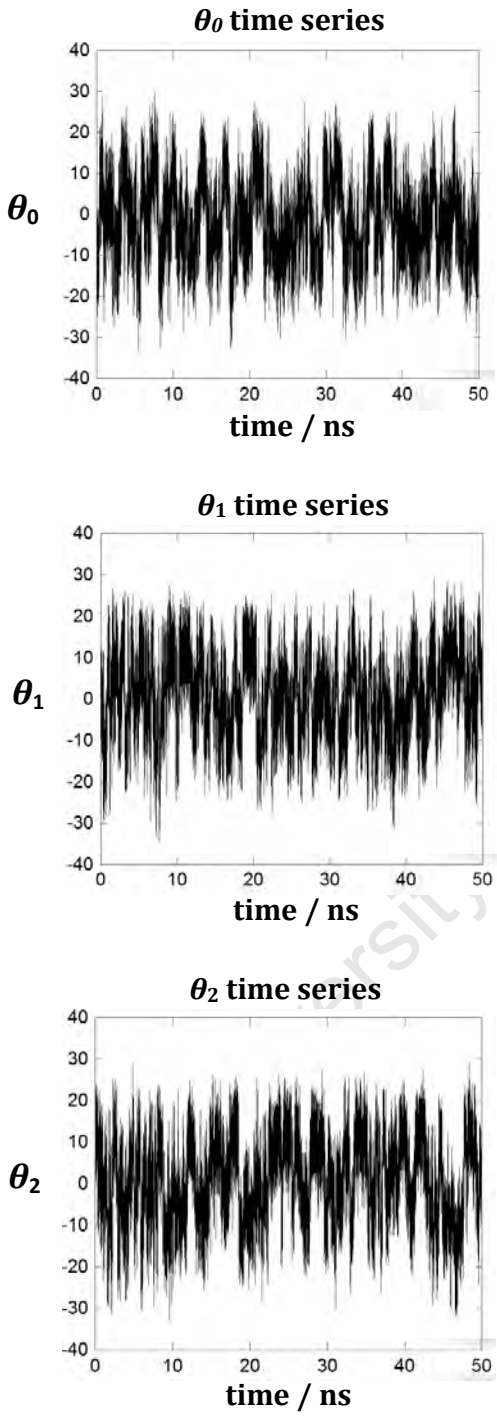


Figure D1: Angular puckering coordinates (APC) time series of α-CD, obtained from the 50 ns MD trajectory. The APC are based on a 6-membered monomeric ring defined by the centres of mass of the CD dimer units, as depicted in **Figure 6.7** of Chapter 6.

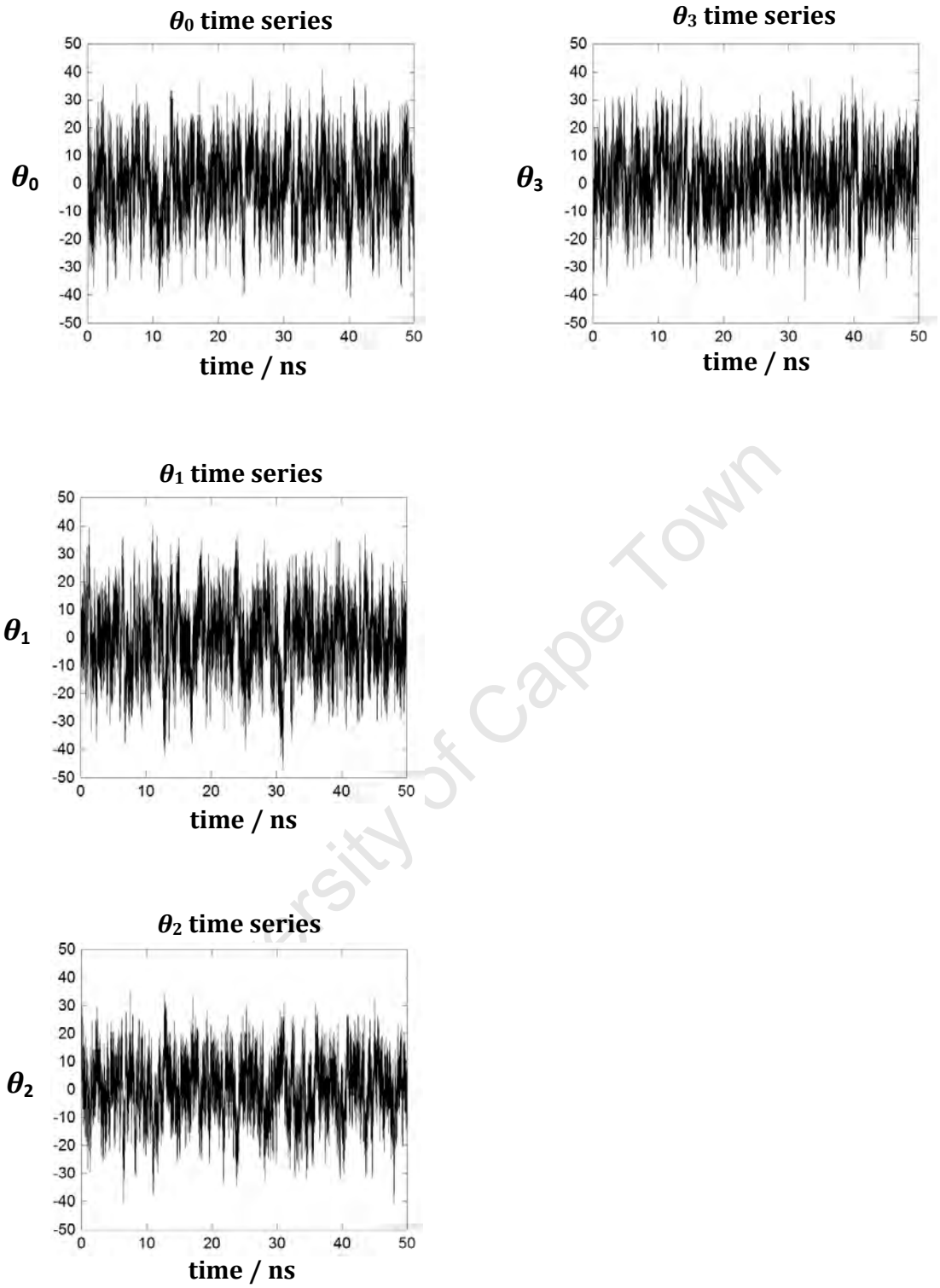


Figure D2: APC time series of β-CD, obtained from the 50 ns MD trajectory. The APC are based on a 7-membered monomeric ring defined by the centres of mass of the CD dimer units, as depicted in **Figure 6.7** of Chapter 6.

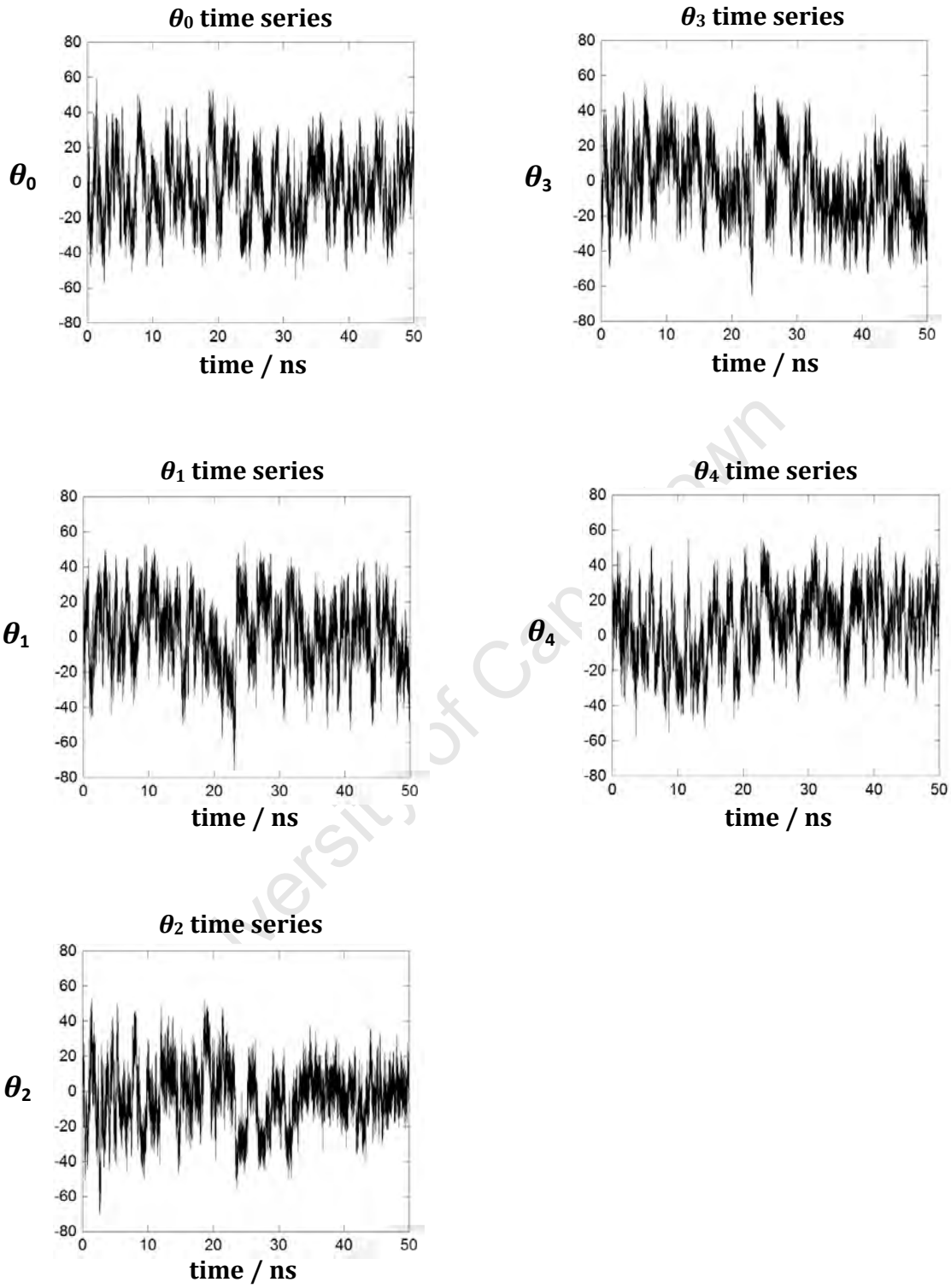


Figure D3: APC time series of γ-CD, obtained from the 50 ns MD trajectory. The APC are based on an 8-membered monomeric ring defined by the centres of mass of the CD dimer units, as depicted in **Figure 6.7** of Chapter 6.